

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

UM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 15:55:18 (Search time 49.1 seconds

(without alignments)  
4030.649 Million cell updates/sec

Title: US-08-799-910-9  
Perfect score: 1228  
Sequence: 1 ATGTCTCAGCTCGCAGCTG.....AAAAAAAAAACTGAG 1228

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/prodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/lna/6.COMB.seq:\*  
4: /cgn2\_6/prodata/2/lna/PCTUS.COMB.seq:\*  
5: /cgn2\_6/prodata/2/lna/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1228	100.0	1228	3	US-08-826-246-9
2	1228	100.0	1228	3	US-08-944-495-9
3	1228	100.0	1228	3	US-09-126-640-5
4	49.8	4.1	7218	1	US-08-232-463-14
5	43.8	3.6	7218	1	US-08-232-463-14
6	42	3.4	8438	1	US-07-945-283-1
7	41.2	3.4	530	2	US-08-758-662-4
8	41.2	3.4	4524	2	US-08-845-998-7
9	41.2	3.4	4524	2	US-09-206-537-7
10	40.6	3.3	15231	3	US-09-128-155-16
11	40	3.3	2847	3	US-08-087-007-2
12	40	3.3	2847	3	US-08-483-433-2
13	40	3.3	2847	4	PCT-US92-05920-2
14	39.8	3.2	1120	3	US-09-030-613-1
15	38.6	3.1	6354	3	US-09-058-389A-5
16	38.4	3.1	6803	3	US-08-665-259-19
17	38.4	3.1	6803	3	US-08-762-500-19
18	37.6	3.1	170	2	US-09-058-389A-13
19	37.6	3.1	4425	2	US-08-749-169A-1
20	37.6	3.1	4425	2	US-09-130-032A-1
21	37.4	3.0	1420	2	US-08-909-965C-3
22	37.2	3.0	371	1	US-08-664-596B-25
23	37.2	3.0	371	2	US-08-739-775-3
24	37.2	3.0	20235	3	US-07-642-734C-3
25	37.2	3.0	20235	3	US-08-439-009A-3
26	36.6	3.0	1931	2	US-09-130-114-2
27	36.4	3.0	702	1	US-08-458-568A-3
28	36.4	3.0	1841	2	US-08-820-521-1

## ALIGNMENTS

```

RESULT 1
US-08-826-246-9
Sequence 9, Application US/08826246
Patent No. 6048709
GENERAL INFORMATION:
APPLICANT: Faib, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSER: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/826,246
FILING DATE: 28-MAR-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/799,910
FILING DATE: 13-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/011,787
FILING DATE: 16-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-078-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1228 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1..468
OTHER INFORMATION:

```





421 tacgtctgagactcagcacttctccagcaacacccgagccttctaaatgtgactc 480  
QY 401 CCCGCACTCCCAAAAGAAATCCGAAAACACAAAGAACCCAGGCTACCTGGTCG 540  
Db 481 cccgcactcccaaaaagaalccgaaaaaacacaaagaacacacagcgatcggcgg 540  
QY 541 CGAGAGCGCTATCCCACTGGGACTTCCGAGCGCAACTTGACACTGACACTACAGCCGA 600  
Db 541 cgaagacglatcccaactcggactccgaagcaactcgaacacacacacacga 600  
QY 601 GAGCGACCCGCTGCTTACGCGGAGCGGACGACGACGACGACGACGACGACGACG 660  
Db 601 gacgcacccggtcttgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 660  
QY 661 GAGCGACGACCCGCTGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 720  
Db 661 gagagacagccagcctgggagcgagcgagcgagcgagcgagcgagcgagcgagc 720  
QY 721 TATTCCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 780  
Db 721 tatgtctcctaaatlaattatlaattatlaattatlaattatlaattatlaatt 780  
QY 781 GTACGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 840  
Db 781 gtaagtaataatlaattatlaattatlaattatlaattatlaattatlaattat 840  
QY 841 AGGTCTCTTGTATTTATTTAGACTTTGTGAGCTGTTGGAAGACGACGACGACGAC 900  
Db 841 aggtctcttgtatltatltatltatltatltatltatltatltatltatltatlt 900  
QY 901 GCGCAAGTACGAGCAAAATCGGAGGAGCTCGGCTGGGAGGAGGAGGAGGAGGAG 960  
Db 901 ggcgaagtagaggaagaatgggagagactcgggagagagagagagagagagag 960  
QY 961 AGGTGTGGTGGGCTCGAAGTTAGAGAGTGTGATGCTGCTGAGTCACTGACGTCG 1020  
Db 961 aaggtgtgggtgggctcgaaagttagagagtgatgctgctgagtcacatccacac 1020  
QY 1021 TCTGTCTACTGTGTGAGACTTGGCGGACCATTAAGAAATGAGATCGTGAGATCTT 1080  
Db 1021 tctgtctactgtgtgagacttggcggaacataggaatgagatccgtagatccctca 1080  
QY 1081 TCTTCTTACAGTCGCTTTAGGCTGGCTCGGACGTAAGAGCTTGGCGCTGCTGCT 1140  
Db 1081 tcttcttgaagtcgacttgaagtgagtgagtgagtgagtgagtgagtgagtgag 1140  
QY 1141 CACGAGGACGCTGAGAGTGGCTAGTGTGTTCTGTGACACAAATAAATTGATTTC 1200  
Db 1141 cagcgaagcagctgctgagagtcgctcagatgttctgtgacacaaataaattgtac 1200  
QY 1201 TGTCAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1228  
Db 1201 tgtcaaaaaataaataaataaataaataaataaataaataaataaataaataa 1228

RESULT 4  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367

GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHREIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)883-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZpL-F15  
US-08-232-463-14

Query Match 4.1% Score 49.8 DB 1: Length 7218;  
Best Local Similarity 1.6% Pred. No. 0.0012;  
Matches 6; Conservative 220; Mismatches 147; Indels 0; Gaps 0;

QY 97 CCGTACATCTTACCTTGACCCCTCTCCGAGCCCGGAGCCCGCCCTCCGCGCCCGCC 156  
Db 1059 CTTGGAGATTT 1118  
QY 157 AGCGCTCTCGCGGACGCAAGAGCGAGCGGAGGCTTCTACCTCGAGTGCTCGG 216  
Db 1119 TT 1178  
QY 217 GCGCAGCTCGAGTGCAGAACCGAACCGAACCGAAAGGCTTCTTCTGCTGAC 276  
Db 1179 TT 1238  
QY 277 ATCGTCTTCCGAGATCTGATGGCTGAGAGGCTGCGCGCGCCCTCCCTCCAG 336  
Db 1239 TT 1298  
QY 337 GAGCGCCCTACGCGGACGACCTTGGCGGACGCGGAGGCTTCTGCTGAGCCCTT 396  
Db 1299 TT 1358  
QY 397 AATCTGACTGCGAGCCCTCGAGACTAGCTGAGACTGAGACTTCTCCAGCAACAC 456  
Db 1359 TT 1418  
QY 457 CCGCGCGCTCT 469  
Db 1419 TTTTTTTTTTTTT 1431

RESULT 5  
US-08-232-463-14/C  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHREIFLINGER, F.  
APPLICANT: FALKNER, F. G.









Query Match	3.3%	Score 40:	DB 1:	Length 2847:
Best Local Similarity	51.1%	Pred. No. 0.3:		
Matches 94:	Conservative 0:	Mismatches 90:	Indels 0:	Gaps 0
Qy	326	TGCGTCAGAGACGCCCTTAACGCCGATCCCTTGCGCCCAACCCCTGTCGCCCTCC	385	
Db	505	TGAATCCAGAGGATGTTCCAGAGACTCTCTCTCTCCCTCCCTCCCACTCTCCCGAGTC	564	
Qy	386	TCGAGCCCTTTAATCTGATCTCGAGCGCCTCGAGCTTAAGCTTTGAGACCTCAGACTTTC	445	
Db	565	TAGCGCCCCGGCGTATGCAGCCCGGAGCCCTTGAGACGCACTCTTATCCACAACAACCC	624	
Qy	446	TCCAGCAACACCGCGCCGCTTCTAACTGTGACTCCCGCACTCCCAAAAAGAAATCCGA	505	
Db	625	TACTCCACCCCTTGTGTTGTCCACCTTGTGTAGCAGAGACCCAGCCAGACCCGCC	684	
Qy	506	AAAA 509		
Db	685	CCAA 688		

RESULT 12  
 US-08-483-433-2  
 ; Sequence 2, Application US/08483433  
 ; Patent No. 6100443  
 GENERAL INFORMATION:

[illegible]

OY 506 AAAA 509  
DB 685 CCAA 688

## RESULT 13

PCT-US92-05920-2

Sequence 2, Application PC/TUS9205920

GENERAL INFORMATION:

APPLICANT: Sims, Peter J.

APPLICANT: Botchwell, Alfred L.M.

APPLICANT: Elliott, Eileen A.

APPLICANT: Flavell, Richard A.

APPLICANT: Madri, Joseph

APPLICANT: Rollins, Scott

APPLICANT: Bell, Leonard

APPLICANT: Squinto, Stephen

TITLE OF INVENTION: Universal Donor Cells

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kilpatrick &amp; Cody

STREET: 1100 Peachtree Street, Suite 2800

CITY: Atlanta

STATE: Georgia

COUNTRY: U.S.

ZIP: 30309-4530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/05920

FILING DATE: 19920714

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Pabst, Patrea L.

REGISTRATION NUMBER: 31,284

REFERENCE/DOCKET NUMBER: OMRF135

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-815-6500

TELEFAX: 404-815-6555

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2847 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

LIBRARY: Genbank HUMDAF: HUMDAF1

CLONE: Human DAF cDNA

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 1..819

OTHER INFORMATION: /note="HUMDAF1 (Promotor and 5'

OTHER INFORMATION: end of Exon 1, genomic sequence)"

PCT-US92-05920-2

Query Match 3.38; Score 40; DB 4; Length 2847;

Best Local Similarity 51.18; Pred. No. 0.3;

Matches 94; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

OY 326 TGCCCTCCAGACAGCCCTTAACGCGCATCCCTGGCCGCCACCCCTGTGTCGCCCGTC 385

DB 505 TGACTCCAGAGCGCTGTCCAGAGAGCTCTCTCCCTCCCTCCCGACCTTCGCCGAGTC 564

OY 386 TCGAGCCCTTAATCTGACTGAGAGCCCTCGACTACGCTCTGACCTCAGCACTTCC 445

DB 565 TAGGGCCCCGGGGTATGAGCGCGGAGCCCTCTGAGCGACCTCTACCAACCAACACCC 624

OY 446 TCCAGACACACCGCGCCCTTCTACTGTGACTCCCGACACTCCCAAAAGAAATCCGA 505

DB 625 TACTCCACCGCTGTGTTGTCACACCTTGSTGACGAGAGCCCGACCGACCGCCGCG 684

OY 506 AAAA 509

DB 685 CCAA 688

## RESULT 14

US-09-030-613-1/C

Sequence 1, Application US/09030613

Patent No. 6083706

GENERAL INFORMATION:

APPLICANT: Florjanczyk, Robert Z.

APPLICANT: Baird, J. Andrew

TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT

NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/030,613

FILING DATE: 25-FEB-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: No. 6083706tenburg Ph.D., Carol

REGISTRATION NUMBER: 39,317

REFERENCE/DOCKET NUMBER: 760100.418C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1120 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-030-613-1

Query Match 3.28; Score 39.8; DB 3; Length 1120;

Best Local Similarity 51.48; Pred. No. 0.22; Mismatches 87; Indels 0; Gaps 0;

Matches 92; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

OY 39 CCTGAGGCCCCGAGCCCCCTTCACCATCCGCGGAGCCGCGGCGCTCCGTC 98

DB 362 CCTGCGGCGCCGCGCGGATCCCGAGCTGTGAGCGCGCGGAGCGCGCGCTCC 303

OY 99 TGAGATCTTACCTTCAGACCTCTCCGCGAGCCCGACGAGCCCTGCGGCGCGCCAG 158

DB 302 CCGGCCCCGCGCGCGCTTCGAGCCGCTCCGCGGAGCGCGCGCGCGCCCGCCAG 243

OY 159 CGCCTCTGCGGCGACCAAGCGAGCGGATCTTACCTTCGAGTGTGTCGCG 217

DB 242 CCTTCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 184

RESULT 15

US-09-058-389A-5

Sequence 5, Application US/09058389A

Patent No. 6130065  
GENERAL INFORMATION:  
APPLICANT: Belt, Judith A.  
APPLICANT: Crawford, Charles R.  
APPLICANT: Patel, Divyen  
TITLE OF INVENTION: A NITROBENZYL MERCAPTOPYRIMIDINE  
TITLE OF INVENTION: (NMPR)-INSENSITIVE, EQUILIBRATIVE, NUCLEOSIDE TRANSPORT  
TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF  
TITLE OF INVENTION: USE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
STREET: Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/058,389A  
FILING DATE: April 9, 1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 1340-1-013N  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ. ID NO.: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6354 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
US-09-058-389A-5

Query Match 3.1%; Score 38.6; DB 3; Length 6354;  
Best Local Similarity 65.9%; Pred. No. 1;  
Matches 56; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
QY 310 GGTGTGCGGCGCCCTGCTCAGAGAGACGCCCTAAGCGCGATCCTGGCGCCAC 369  
DB 2461 GGTGAGACCCCTGCGCGCTGCGCGCTGCGCGCTGCGCGCGACCTTATGAGGCCCTCC 2520  
QY 370 CCTGTGTCCTCCGCTCTGAGGCCT 394  
DB 2521 CCTGCGCCCTCTGCGCTCAGGCCT 2545

Search completed: January 30, 2001, 21:28:30  
Job Time: 19992 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 16:08:34 ; Search time 77.57 Seconds  
(without alignments)  
5947.064 Million cell updates/sec

Title: US-08-799-910-9  
Perfect score: 1228  
Sequence: 1 ATCTGTCACCTCTGCAGCTG.....AAAAAAAAAACTCGAC 1228

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapect 1.0

Searched: 48022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

N\_Geneseq\_36:\*

- 1: /cgn2\_2/gcgdata/geneseq/geneseq/NA1980.DAT:\*
- 2: /cgn2\_2/gcgdata/geneseq/geneseq/NA1981.DAT:\*
- 3: /cgn2\_2/gcgdata/geneseq/geneseq/NA1982.DAT:\*
- 4: /cgn2\_2/gcgdata/geneseq/geneseq/NA1983.DAT:\*
- 5: /cgn2\_2/gcgdata/geneseq/geneseq/NA1984.DAT:\*
- 6: /cgn2\_2/gcgdata/geneseq/geneseq/NA1985.DAT:\*
- 7: /cgn2\_2/gcgdata/geneseq/geneseq/NA1986.DAT:\*
- 8: /cgn2\_2/gcgdata/geneseq/geneseq/NA1987.DAT:\*
- 9: /cgn2\_2/gcgdata/geneseq/geneseq/NA1988.DAT:\*
- 10: /cgn2\_2/gcgdata/geneseq/geneseq/NA1989.DAT:\*
- 11: /cgn2\_2/gcgdata/geneseq/geneseq/NA1990.DAT:\*
- 12: /cgn2\_2/gcgdata/geneseq/geneseq/NA1991.DAT:\*
- 13: /cgn2\_2/gcgdata/geneseq/geneseq/NA1992.DAT:\*
- 14: /cgn2\_2/gcgdata/geneseq/geneseq/NA1993.DAT:\*
- 15: /cgn2\_2/gcgdata/geneseq/geneseq/NA1994.DAT:\*
- 16: /cgn2\_2/gcgdata/geneseq/geneseq/NA1995.DAT:\*
- 17: /cgn2\_2/gcgdata/geneseq/geneseq/NA1996.DAT:\*
- 18: /cgn2\_2/gcgdata/geneseq/geneseq/NA1997.DAT:\*
- 19: /cgn2\_2/gcgdata/geneseq/geneseq/NA1998.DAT:\*
- 20: /cgn2\_2/gcgdata/geneseq/geneseq/NA1999.DAT:\*
- 21: /cgn2\_2/gcgdata/geneseq/geneseq/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1228	100.0	1228	21	250711
2	1226.4	99.9	1228	18	794471
3	420.8	34.3	427	21	280483
4	371.6	30.3	411	20	V86237
5	167.2	13.6	193	20	V89271
6	151	12.3	297	15	Q77534
7	70.4	5.7	267	15	Q77534
8	57.4	4.7	1000	21	A02484
9	55.2	4.5	1218	21	A02488
10	52.2	4.3	1337	20	217253
11	49.6	4.0	1126	21	A02538
12	48.2	3.9	53	16	T20789

13	47.8	3.9	4228	21	251683
14	47.2	3.8	114955	20	X53491
15	44.8	3.6	1593	21	A02504
16	44.2	3.6	2375	21	A34740
17	44.2	3.6	4475	21	A34741
18	42.8	3.5	1127	21	A02477
19	42.4	3.5	4356	14	Q37543
20	42.4	3.5	4356	16	Q95540
21	42	3.4	8438	15	Q73500
22	42	3.4	34094	20	230163
23	41.8	3.4	1459	21	A02528
24	41.6	3.4	989	21	A02539
25	41.4	3.4	1275	11	Q01620
26	41.4	3.4	1344	12	Q13318
27	41.4	3.4	1345	11	Q03062
28	41.2	3.4	1102	20	216923
29	41.2	3.4	4524	20	V33912
30	41	3.3	114955	20	X53491
31	40.8	3.3	1966	20	X61220
32	40.6	3.3	1017	20	217058
33	40.4	3.3	300	20	214922
34	40.4	3.3	2179	15	Q73117
35	40.4	3.3	2232	15	Q77739
36	40	3.3	1523	20	217507
37	40	3.3	1523	20	217479
38	39.8	3.2	1120	19	V60339
39	39.8	3.2	2188	20	277506
40	39.8	3.2	3198	20	X02974
41	39.6	3.2	954	20	X30342
42	39.6	3.2	2681	20	X58429
43	39.2	3.2	1804	20	X61218
44	38.8	3.2	749	20	216153
45	38.8	3.2	1312	20	217264

#### ALIGNMENTS

RESULT 1	
ID 250711	standard; DNA: 1228 BP.
AC 250711;	
DT 31-MAY-2000	(first entry)
DE	Nucleotide sequence of human fchd605 gene.
XX	
KW	fchd605 gene; human; cardiovascular disease; oncogenic disorder;
KW	diabetic retinopathy; fibroproliferative disorder; arteriosclerosis;
KW	TGF-beta signalling pathway; TGF; Transforming growth factor;
KW	pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;
KW	vascularisation; cytoskeletal; antidiabetic; ophthalmological; ds.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Location/Qualifiers
FT	1..471
FT	CDS
FT	/tag= a
FT	/product= "fchd605 protein"
XX	
PN	WO200006206-A1.
XX	
PD	10-FEB-2000.
XX	
PR	30-JUL-1999; 99WO-US17394.
XX	
PR	30-JUL-1998; 98US-0126640.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Fold DA:
XX	

Human cyclic nucle  
Human adenosine A1  
Human colon cancer  
Human adenosine re  
Human adenosine re  
Human colon cancer  
Cardiac adenylyl c  
Cardiac adenylyl c  
DNA encoding Pseud  
Complete nucleotid  
Human colon cancer  
Human colon cancer  
Chicken beta-actin  
Expression vector  
Hybrid promoter of  
Human gene express  
Nucleotide sequenc  
Human adenosine A1  
Mouse DNA demethyl  
Human gene express  
Human gene express  
Alpha 2, 3-sialyl  
Human alpha2,3-sl  
Human gene express  
Human gene express  
CDNA sequence of f  
Human ovarian luno  
Human IL-1ra BAC c  
DNA encoding a hum  
Thermophilus therm  
Human DNA demethyl  
Human gene express  
Human gene express

DK WP1: 2000-205414/18.  
 DR P-PSDB: Y45017.

PT Identifying substances for ameliorating symptoms of fibroproliferative  
 diseases or oncogenic related disorders -

XX Examples; Fig 5; 214pp; English.

CC The patent discloses methods for the treatment and diagnosis of  
 CC cardiovascular diseases by novel human genes which are differentially  
 CC expressed in different cardiovascular disease states. Compositions which  
 CC can modify TGF-beta signalling pathway are identified by screening.  
 CC These are used therapeutically to treat fibroproliferative and oncogenic  
 CC disorders, especially TGF (Transforming growth factor)-beta related  
 CC disorders, including diabetic retinopathy, arteriosclerosis, pancreatic  
 CC cancer, angiogenesis, inflammation, fibrosis, tumour growth and  
 CC vasculature. The present sequence is fchd605 gene which is up  
 CC -regulated in monocytes treated with oxidised LDL (low density  
 CC lipoprotein) can be used to design cardiovascular disease treatment  
 CC strategies. Depending on whether the up-regulation has a pathogenic or  
 CC protective effect treatment methods can be designed to increase or  
 CC decrease the activity of the protein product of the gene.

XX Sequence 1228 BP: 265 A; 361 C; 341 G; 261 T; 0 other:

Query Match 100.0%; Score 1228; DB 21; Length 1228;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATCTGTCATCTTCGACGCTGCGACCGACCGACCATCTGCGAGCCCGGCGCGCC 60  
 DB 1 ATCTGTCATCTTCGACGCTGCGACCGACCGACCGACCATCTGCGAGCCCGGCGCGCC 60  
 OY 61 CCTTCACCATCCCGGAGCCCGGCGCGCGCTCGGTCTGAGATCTTCACTTCGACCT 120  
 DB 61 CCTTCACCATCCCGGAGCCCGGCGCGCGCTCGGTCTGAGATCTTCACTTCGACCT 120  
 OY 121 CTCTCCGAGACCGCGACCGCCCTGCGCGCGCCCGACGCGCTCTGCGCGACCGAAG 180  
 DB 121 CTCTCCGAGACCGCGACCGCCCTGCGCGCGCCCGACGCGCTCTGCGCGACCGAAG 180  
 OY 121 CTCTCCGAGACCGCGACCGCCCTGCGCGCGCCCGACGCGCTCTGCGCGACCGAAG 180  
 DB 121 CTCTCCGAGACCGCGACCGCCCTGCGCGCGCCCGACGCGCTCTGCGCGACCGAAG 180  
 OY 181 CGGAGCGCGAGGCTTCTACCTCGAGTGTGCGCGCGCGACGCTGCGACGACCG 240  
 DB 181 CGGAGCGCGAGGCTTCTACCTCGAGTGTGCGCGCGCGACGCTGCGACGACCG 240  
 OY 181 CGGAGCGCGAGGCTTCTACCTCGAGTGTGCGCGCGCGACGCTGCGACGACCG 240  
 DB 181 CGGAGCGCGAGGCTTCTACCTCGAGTGTGCGCGCGCGACGCTGCGACGACCG 240  
 OY 241 AACCCAGCCAAAGGCTTCTTCTTCTGCTGCTACCATGCTTCTGCGACATCTGATG 300  
 DB 241 AACCCAGCCAAAGGCTTCTTCTTCTGCTGCTACCATGCTTCTGCGACATCTGATG 300  
 OY 241 AACCCAGCCAAAGGCTTCTTCTTCTGCTGCTACCATGCTTCTGCGACATCTGATG 300  
 DB 241 AACCCAGCCAAAGGCTTCTTCTTCTGCTGCTACCATGCTTCTGCGACATCTGATG 300  
 OY 301 GCTGAAGAGGCTGTGCGCGCGCGCTGCTGCGAGAGCGCGCTTAACGCGCATGCGTG 360  
 DB 301 GCTGAAGAGGCTGTGCGCGCGCGCTGCTGCGAGAGCGCGCTTAACGCGCATGCGTG 360  
 OY 301 GCTGAAGAGGCTGTGCGCGCGCGCTGCTGCGAGAGCGCGCTTAACGCGCATGCGTG 360  
 DB 301 GCTGAAGAGGCTGTGCGCGCGCGCTGCTGCGAGAGCGCGCTTAACGCGCATGCGTG 360  
 OY 361 GCGCGCACCGCTGCTGCGCGCGCGCTGCGAGCGCTTAATCTGACTTCGAGCGCGAG 420  
 DB 361 GCGCGCACCGCTGCTGCGCGCGCGCTGCGAGCGCTTAATCTGACTTCGAGCGCGAG 420  
 OY 361 GCGCGCACCGCTGCTGCGCGCGCGCTGCGAGCGCTTAATCTGACTTCGAGCGCGAG 420  
 DB 361 GCGCGCACCGCTGCTGCGCGCGCGCTGCGAGCGCTTAATCTGACTTCGAGCGCGAG 420  
 OY 421 TAGCTCTGAGACTTCAGACATTTCTCTCAGACAAACCGCGCGCTTCTAATCTGACTC 480  
 DB 421 TAGCTCTGAGACTTCAGACATTTCTCTCAGACAAACCGCGCGCTTCTAATCTGACTC 480  
 OY 421 TAGCTCTGAGACTTCAGACATTTCTCTCAGACAAACCGCGCGCTTCTAATCTGACTC 480  
 DB 421 TAGCTCTGAGACTTCAGACATTTCTCTCAGACAAACCGCGCGCTTCTAATCTGACTC 480  
 OY 481 CCGCGCATTCGCGCAAAAGATCGAAGAAACCAAGAAACCAAGCGGCTGCTGCTGCG 540  
 DB 481 CCGCGCATTCGCGCAAAAGATCGAAGAAACCAAGAAACCAAGCGGCTGCTGCTGCG 540  
 OY 481 CCGCGCATTCGCGCAAAAGATCGAAGAAACCAAGAAACCAAGCGGCTGCTGCTGCG 540  
 DB 481 CCGCGCATTCGCGCAAAAGATCGAAGAAACCAAGAAACCAAGCGGCTGCTGCTGCG 540  
 OY 541 CGAGACCGCTATCCCACTGCGACTTCGAGCGCAACTTGAATCTGAGACATTCAGCGCA 600  
 DB 541 CGAGACCGCTATCCCACTGCGACTTCGAGCGCAACTTGAATCTGAGACATTCAGCGCA 600  
 OY 541 CGAGACCGCTATCCCACTGCGACTTCGAGCGCAACTTGAATCTGAGACATTCAGCGCA 600  
 DB 541 CGAGACCGCTATCCCACTGCGACTTCGAGCGCAACTTGAATCTGAGACATTCAGCGCA 600  
 OY 601 GAGCGCACCGCGCTGCTGAGCGCGGAGCGCGAGCGCGACAGAGCGCGCGCATAGAGACG 660  
 DB 601 GAGCGCACCGCGCTGCTGAGCGCGGAGCGCGAGCGCGACAGAGCGCGCGCATAGAGACG 660  
 OY 601 GAGCGCACCGCGCTGCTGAGCGCGGAGCGCGAGCGCGACAGAGCGCGCGCATAGAGACG 660  
 DB 601 GAGCGCACCGCGCTGCTGAGCGCGGAGCGCGAGCGCGACAGAGCGCGCGCATAGAGACG 660

OY 661 GAGCGCACCGCGCTGCTGAGCGCGGAGCGCGAGCGCGACAGAGCGCGCGCATAGAGACG 720  
 DB 661 GAGCGCACCGCGCTGCTGAGCGCGGAGCGCGAGCGCGACAGAGCGCGCGCATAGAGACG 720  
 OY 721 TATTCGCTGCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 780  
 DB 721 TATTCGCTGCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 780  
 OY 781 GTACGTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 840  
 DB 781 GTACGTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 840  
 OY 841 AGCTCTGCTGCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 900  
 DB 841 AGCTCTGCTGCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 900  
 OY 901 GCGCAATAG 960  
 DB 901 GCGCAATAG 960  
 OY 961 AAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
 DB 961 AAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
 OY 1021 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080  
 DB 1021 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080  
 OY 1081 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140  
 DB 1081 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140  
 OY 1141 CAGCGAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
 DB 1141 CAGCGAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
 OY 1201 TGTCAAAAAAAG 1260  
 DB 1201 TGTCAAAAAAAG 1260  
 RESULT 2  
 T94471  
 ID T94471 standard; cDNA: 1228 BP.  
 XX T94471:  
 NC T94471:  
 DP 03-MAR-1998 (first entry)  
 XX  
 DE Human Fchd605 gene differentially regulated in monocytes.  
 KW Fchd605 gene; differential expression; monocyte; human;  
 KW foam cell; cardiovascular disease; arteriosclerosis; ischemia;  
 KW reperfusion; hypertension; restenosis; arterial inflammation;  
 KW therapy; diagnosis; drug screening; marker; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 1..471  
 FT tag a  
 PD M09730065-A1.  
 XX 21-AUG-1997.  
 XX 14-FEB-1997; 97WO-US02291.  
 XX 13-FEB-1997; 97US-0799910.  
 XX 16-FEB-1996; 96US-0011787.  
 XX (MILL-) MILLENNIUM PHARM INC.



XX Falb DA;  
PI  
XX  
DR WPT: 1997-424966/39.  
DK P-PSDB; W36006.

PT New genes differentially expressed in cardiovascular disease - used  
PT for diagnosis, drug screening and treatment of cardiovascular  
PT disease, e.g. atherosclerosis, restenosis, hypertension, etc  
XX  
PS Claim 1; Fig 5; 163pp: English.

Claim 1; Fig 5; 163pp; English.

Fchb605 is a novel human gene that is up-regulated in monocytes treated with oxidised low-density lipoproteins (LDL). Differential display was used to detect genes that are differentially expressed in monocytes treated so as to simulate the conditions under which foam cells develop during atherogenesis. Both fchb605 and fchb602 (see T94470) are up-regulated under the disease condition of treatment with oxidised LDL. The fchb605 gene product (see W36006) has sequence similarity to mouse gly9g. The discovery of the up-regulation of these 2 genes provides a fingerprint profile, e.g. markers for the study of cardiovascular diseases, including atherosclerosis, ischemia/reperfusion, hypertension, restenosis, and arterial inflammation. Methods are provided for the diagnosis, monitoring in clinical trials, screening for therapeutically effective compounds, and treatment of cardiovascular diseases based on discoveries regarding the expression patterns of novel genes fchb531 (see T94467), fchb540 (see T94468), fchb545 (see T94469), fchb502 and fchb605.

Sequence 1228 BP; 265 A; 362 C; 340 G; 261 T; 0 other.

query match	99.98;	Score 1226.4;	DB 18;	Length 1228
-------------	--------	---------------	--------	-------------

```
Matches 1227; Conservative 0; Mismatches 1; Indels 0; Gaps 0.
```

[illegible]

OY		541	CGAGAGGATTTCCCAACTGGGACCTTCGAGGCACACTTGAACTCAGAACACTACAGUGA	600
Db		541	cgaagaggatctccccaactcgggaccttcgaggaaccaattcgaaacactcaagggga	600
OY		601	GAGGCCACC CGGTGCTTGAAGCGCGGACCGAGGCGGACAGAGACCGAGGCCATTCAGACC	660
Db		601	gagccaccaccggtgccttgaagcgggaccgagcgcacagaagaccgagcgcalaaagacc	660
OY		661	GAGGCACAGCCCAGACTGGGGCTTAGGCCCGGTGGGAAGCAGACCGCTTAATTATTTC	720
Db		661	gaggacaagcccacgctggggcttaggcccggtgggaagagagtgctgtaattattctc	720
OY		721	TATTGCTCCCTAAATAATTATTATATATATATTATATGACGCTCCTAGGTGATGGAGATG	780
Db		721	tattgctccctaataataattatatatgatatttatgctacgcctccaaagtgtgagagatg	780
OY		781	GTACGTAATATTTATTTTTAACTTATACCAAGGCTGTGACATGTTCCTCTGCTTAATTC	840
Db		781	gtacgtaaatattttattttaacttatgcgaagggtgtagatgtlccctctgcgtlaaaatgc	840
OY		841	AAGTCTCTGTGTATTTATTGAGCTTTGTGGGACTGTGTGAACACAGACACTGCAATGC	900
Db		841	aagtcctctgtgtattttattgagctttgtgggactgtgtgaacacagacactgcaatgc	900
OY		901	GGCAAAGTAGAGAAGAATGGGAGGACTCGGCTGGGAGGACGTCGCCGCTTGGATG	960
Db		901	ggcaaagtagagaagaatgggaggactcggctgggaggacgtcccgcttggatg	960
OY		961	AAGTCTGTGCTGGCTGGCTGTAAGTTTAGAGAGTACAGTCCCTCCAGATCTCACATCCG	1020
Db		961	aagtcctgtgctggctggctgtaagtttagagagtacagtccctccagatctcacatccg	1020
OY		1021	TCTGTCTACTGTGTGAGACTTTCGCGGACCATTAGGAATGAGATCCGTGAGATCTCTCCA	1080
Db		1021	tctgtctactgtgtgagactttcgcggaccatttaggaatgagatccgtgagatctctcca	1080
OY		1081	TCTTTTAAAGTCGCCCTTTAGAGTGGCTGCGGAGGTAGAGGCTTGGGCGTGGCTGT	1140
Db		1081	tctttttaaagtcgcccttttagagtggtcgcggaggtagaggcttgggcggtggctgt	1140
OY		1141	CACGAGCAGCATGTGCGAGATGCGCTACTATGTTCTGTGAACACAAAATAAATTTGATTTAC	1200
Db		1141	cacgagcagcatgtgcgagatgcgctactatgttctgtgaacacaaaataaattgatttac	1200
OY		1201	TGTCTAAAAAATAAAAAAAAAAACTCGAG	1228
Db		1201	tgtctaaaaaataaaaaaaaaaac t cgag	1228
<hr/>				
RESULT	3			
280483/C				
ID	280483	standard;	cdNA; 427 BP.	
XX	XX	280483;		
DT	07-APR-2000	(first entry)		
De	Human colon cancer cell line SW480	CDNA clone SEQ ID NO:567.		
XX	XX	Human: gene expression product; diagnosis; tumour; colon cancer;		
KW	XX	colorectal adenocarcinoma; cell line SW480; cell proliferation;		
KM	XX	cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;		
KX	XX	hyperplasia; ds.		
OS	XX	Homo sapiens.		
PN	XX	WO964576-A2.		
PD	XX	16-DEC-1999.		
Pf	XX	09-JUN-1999; 99WO-IB01062.		
PR	XX	10-JUN-1996; 98US-0088801.		



OY 824 CCTCTGTGTAATGACAGTCTCTGATTTATGACCTTTGTGGACTGTGGAGC 883  
 DB 315 cccnctgtctgtaataatgacaggtctctgtgattatgagcttctggagactgttgaagc 374  
 OY 884 AGCAGCCTGGAACCTGCGGCAAGAGAGA 914  
 DB 375 aggaacacctggacatgagcgaagagagcga 405

RESULT 5

ID V89271/C  
 V89271 standard: CDNA: 193 BP.

XX V89271:

XX 15-FEB-1999 (first entry)\*

DE EST clone BY66.

XX Human: secreted protein: expressed sequence tag: EST: haematopoiesis:

KW tissue growth: activin; inhibin; chemokinesis; chemokinesis; haemostatic;

KW receptor; ligand: thrombolytic; anti-inflammatory; cadherin; anti-tumour;

XX gene therapy; ss.

XX Homo sapiens.

XX MO9845436-A2.

XX 15-OCT-1998.

XX 10-APR-1998: 98WO-US06955.

XX 10-APR-1997: 97US-0838821.

XX (GEM) GENETICS INST INC.

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

PI Racie LA, Spaulding V, Treacy M;

XX WPI: 1999-070077/06.

XX New polynucleotides encoding human secreted proteins - derived from

PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,

PT ovary, pituitary, retina and colon CDNA libraries.

XX Claim 1: Page 164: 618bp; English.

XX The present sequence represents a human expressed sequence tag (EST).

XX The polynucleotide, which is a secreted EST, and the encoded protein

CC are predicted to have useful biological activities which would make

CC them suitable for treating, preventing or ameliorating medical

CC conditions in humans and animals, although no supporting data is

CC given. Suggested activities include nutritional activity, immune

CC stimulating or suppressing activity, activin/inhibin activity,

CC chemotactic/chemokinetic activity, haemostatic and thrombolytic

CC activity, receptor/ligand activity, anti-inflammatory activity,

CC cadherin/tumour invasion suppressor activity, tumour inhibition

XX activity. The polynucleotide may also be useful for gene therapy.

XX Sequence 193 BP: 35 A: 39 C: 68 G: 51 T: 0 other:

XX Query Match 13.6%: Score 167.2: DB 20: Length 193:

XX Best local Similarity 98.3%: Pred. No. 4.5e-35:

XX Matches 169: Conservative 0: Mismatches 3: Indels 0: Gaps 0:

OY 385 CTCGAGCCCTTAATGACTTGGAGCCCTGGAGTACGCTCTGACCTCAGACTTC 444

DB 193 CTCGAGCCCTTAATGACTTGGAGCCCTGGAGTACGCTCTGACCTCAGACTTC 134

OY 445 CTCGAGCAGACCCGCGGCTTCTAATCTGACTCCCGAGCTCCCAAGAAATCG 504

DB 133 CTCGAGCAGACCCGCGGCTTCTAATCTGACTCCCGAGCTCCCAAGAAATCG 74  
 OY 505 AAAAACCAGAAAGAACACAGCCGTACCTGTCGCGAGAGCCTATCCCA 556  
 DB 73 AAAAACCAGAAAGAACACAGCCGTACCTGTCGCGAGAGCCTATCCCA 22

RESULT 6

ID Q77534 standard: DNA: 297 BP.

XX Q77534:

XX 23-SEP-1994 (first entry)

DE Human genome fragment.

XX Brain; placenta; bone marrow; genetic analysis: gene mapping;

KW detection; homology: human; adrenal tissue; ds.

XX Homo sapiens.

XX WO9401548-A.

XX 20-JAN-1994.

XX 13-JUL-1993: 93WO-GB01467.

XX 13-JUL-1992: 92GB-0014857.

XX (MEDI-) MEDICAL RES COUNCIL.

XX Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;

PI Sibson DR, Starkey M;

XX WPI: 1994-035056/04.

XX New nucleic acid fragment encoding gene products - can be used

PT for genetic analysis and mapping

XX Claim 1: Page 575-576: 616bp; English.

XX Human nucleic acid fragments, isolated from brain, adrenal tissue,

CC the placenta or bone marrow comprise any of: (A) a sequence

CC selected from (Q76401-Q77613), (B) an allelic variation of a

CC sequence as described in (A), or (C) a sequence complementary

CC to (A) or (B).

CC Preferred sequences exhibit no more than 90% homology to a human

CC sequence known per se.

XX Sequence 297 BP: 66 A: 117 C: 61 G: 52 T: 1 other:

XX Query Match 12.3%: Score 151: DB 15: Length 297:

XX Best local Similarity 76.0%: Pred. No. 1e-30: Indels 2: Gaps 2:

XX Matches 212: Conservative 0: Mismatches 65: Indels 2: Gaps 2:

OY 1 ATGTCGACTCTGCGAGCTGCCAGCCGACCATGACCTCTGAGCCCGGAGCC 60

DB 16 atggaacacatctcgcgcgcgcacatcccaagaccacccaaagcccaagcc 75

OY 61 CCTCCACCATCCCGGAGCCCGGCGGCTCCGCTCTGAGATCTTACCTTCGACCT 120

DB 76 cctcaagctatctcgaagcaataa-gctgcgcgcctgagatctcgaagccct 134

OY 121 CTCGAGAGCCGCGAGCCGCTGCGGAGCCCGGAGCCGCTCTGCGGAGCCGAG 180

DB 135 ctccgaaagcccaagcgcgcctcgcgcgcgcacccagctc-ctcgaaggaacaaag 193

OY 181 CGGAGCCGAGGTTCTTACCTCGAGTGTCCGCGCAGCTGCGCACTGACGAGCC 240

DB 194 cgaagctcgaagatctcgaagcctagagatgacgcgcgcgcagctcgaagcgaagc 253







CC The polynucleotides sequences can be used in a method for detecting  
CC differentially expressed genes correlated with a cancerous state of a  
CC mammalian cell. The polynucleotides can also be used as probes for  
CC detecting and mapping related genes. They can be used in diagnosis and  
CC prognosis of diseases and disorders (e.g. identification of  
CC pre-metastatic or metastatic cancerous states, stages of cancer, or  
CC responsiveness of cancer to therapy). This is particularly for breast  
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-  
CC negative breast cancer, lung cancer, and colon cancer.  
XX  
SQ Sequence 1126 BP; 101 A; 102 C; 575 G; 57 T; 291 other;

Query Match 4.0%; Score 49.6; DB 21; Length 1126;  
Best Local Similarity 33.2%; Pred. No. 0.0012;  
Matches 192; Conservative % 0; Mismatches 383; Indels 4; Gaps 1;  
QY 13 CCACCTGCGCCGACCATGACATCTCTGACGGCCGACCCCGGCCCCCTCCACCATC 72  
DB 608 CCGGNNCCCTCCNNAACNCGCCCTCCCGAGNCGCCATTCCCNCGCCGCTCAAG 549  
QY 73 CCGGACCCCGCGGGCTCGGCTCTGAGATCTTACCTTCGACCTCTCCCGGAGCC 132  
DB 548 CTCGNNCCCCCGGCGCTTCTCCNCCNCCNCCNCCCGCCCGCCCGGCGGCN 489  
QY 133 CGACGGCGCCCTGCGGGCGCCCGACGCGCTCTCGGGGACCCGAAAGCGACGCGAGG 192  
DB 48A CCTGCCNCCNCCCTCCCGCCCGACANGCGGCGGCGCCNCCGCGCCNCG- 428  
QY 193 GTTCTACCTTCGAGTGTGCGCGCGCGAGCTGCGAGTGGAGGACCCAGCCGCAAA 252  
DB 429 ---CCANTCCTGCTCTCTCANCCTGCTTCTCAGCGGTTCCTCTCGNCCNCCNCCCTNCC 373  
QY 253 AGGCTCTCTTCTGCTGCTCACCATCTGCTTCTCCAGATCTCTGATGGCTGAAGGGT 312  
DB 372 CGCTTCNCCCTCTCNCNCCGCTACNCCNCCCGCCGCGCCNCCNCCCTTCGCTGCC 313  
QY 313 CTGCGCGCGCCCTGCTCCAGAGGACGCGCCCTTAACGCGCGATCCTGCGCGCCACCCCT 372  
DB 312 CCGCGCGGCGGCGGCGGCGCTAAAGCCCNANNAAGTNCNCCGCGCCCGCCNCG 253  
QY 373 GTGTCCCGCTCTCGAGCCCTTTAATGACTTTCGGAGCCCTCGGACTACGCTCTGGAC 432  
DB 252 NCNCCGCGCCNCCNCCGCGCTTCCGCGCCNCCNCCGCGCTTCCGCGCGGCTCCNAAA 193  
QY 433 CTCAGCACTTCTCTCAGCAACACGCGCGCTTCTTAATCTGACTTCCCGCACTCCCC 492  
DB 192 ANCANATCNNN 133  
QY 493 AAAAAAATCCGAAAAACACAAAGAACACACAGGCGTACCTGTCGCGGAGAGGCTATC 552  
DB 132 CNTCCGCGCCNCCNCCNCCCGCTTCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 73  
QY 553 CCCAACTGGGACTTTCGAGGCAACTTGAATCAGAACAC 591  
DB 72 NNTCCNCCNCCCTTNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 34

RESULT 12  
T20789  
ID T20789 standard; cDNA to mRNA; 53 BP.  
XX  
AC T20789;  
XX  
DT 09-JUL-1996 (first entry)  
XX  
DE Human gene signature HUMGS02013.  
XX  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
XX cell typing; abnormal cell function; ss.  
OS Homo sapiens.

XX PN W05914772-A1.  
XX  
PD 01-JUN-1995.  
XX  
PF 11-NOV-1994; 94WO-JP01916.  
XX  
PK 12-NOV-1993; 93JP-0355504.  
XX  
PA (MATSU) MATSUBARA K.  
PA (OKUBO) OKUBO K.  
XX  
PI Matsubara K, Okubo K;  
DR WPI; 1995-206931/27.  
XX  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
XX  
PS Claim 1; Page 723; 2245pp; Japanese.  
XX  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared from  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
XX  
SQ Sequence 53 BP; 18 A; 9 C; 9 G; 17 T; 0 other;  
Query Match 3.9%; Score 48.2; DB 16; Length 53;  
Best Local Similarity 94.3%; Pred. No. 0.00069;  
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1158 GATCCCTAGTATGTTCTGTGAACACAAATAAATTGATTACTGTCAAAAA 1210  
DB 1 gatcgctagtgtctgtgaacacaaataaattgatttactgtctgcaaa 53  
RESULT 13  
251683  
ID 251683 standard; cDNA; 4228 BP.  
XX  
AC 251683;  
XX  
DT 04-JUL-2000 (first entry)  
XX  
DE Human cyclic nucleotide-associated protein-2 (CNAP-2) cDNA.  
XX  
KW Cyclic nucleotide-associated protein-2; CNAP-2; human; cytostatic;  
KW anti-arteriosclerotic; hepatotropic; anti-leukaemic; anti-inflammatory;  
KW immunomodulatory; anti-asthmatic; anti-anemic; anti-diabetic; diagnosis;  
KW anti-sclerotic; dermatological; neuroprotective; anti-epileptic; cancer;  
KW anti-Alzheimer's; anti-Parkinsonian; cerebroprotective; ophthalmological;  
KW anti-infertility; anti-allergic; vasotropic; immunosuppressive;  
KW hypotensive; gene therapy; prevention; treatment; arteriosclerosis;  
KW cell proliferative disorder; autoimmune/inflammatory; diabetes mellitus;  
KW neurological; vision; reproductive; smooth muscle; ss.  
XX  
OS Homo sapiens.

FN	Key	Location/Qualifiers	
FT	CDS	31..4014	
FT		/*tag= a	
FT		/product= "Human CNAP-2 protein"	
FT		/note= "Shares 24% identity to Aquifex pyrophilus	
FT		esterase 28LC"	
FT	sig_peptide	31..132	
FT		/*tag= b	
FT	mat_peptide	133..4011	
FT		/*tag= c	
FT		/product= "Mature CNAP-2 protein"	
FT	misc_binding	136..165	
FT		/*tag= d	
FT		/bound_moiety= "Primer or Probe"	
FT		/note= "Useful for amplification or hybridisation	
FT		techniques"	
XX			
FN	W0200014248-Al.		
XX			
XX		16-MAR-2000.	
XX			
XX		03-SEP-1999; 99HO-US20287.	
PF			
XX			
PR		04-SEP-1998; 98US-0148904.	
XX			
PA	(INCY-)	INCYTE PHARM INC.	
XX			
PI	Hillman JL, Yue H, Guegler KJ, Corley NC, Patterson C, Tang YT;		
XX			
DR	WPI; 2000-256994/22.		
DR	P-PSDR; Y70474.		
XX			
PT	Isolated cyclic nucleotide associated proteins useful for preventing,		
PT	diagnosing and treating cell proliferative, autoimmune/inflammatory,		
PT	neurological, vision, reproductive and smooth muscle disorders -		
XX			
PS	Claim 9; Page 70-71; 78pp; English.		
XX			
CC	The present sequence is the cDNA encoding human cyclic nucleotide		
CC	associated protein-2 (CNAP-2), identified in Incyte clone 3149674,		
CC	that is isolated from ADREN004 cDNA library. It is expressed in		
CC	nervous, reproductive, cardiovascular and haematopoietic/immune tissues.		
CC	CNAP sequences may be used for prevention, treatment and diagnosis of		
CC	diseases associated with altered CNAP expression such as, cell		
CC	proliferative disorders (e.g. arteriosclerosis, cirrhosis, leukaemia,		
CC	lymphoma and cancer of the breast, prostate, lung and brain), autoimmune/		
CC	inflammatory disorders (e.g. asthma, anaemia, diabetes mellitus, multiple		
CC	sclerosis and psoriasis), neurological disorders (e.g. epilepsy,		
CC	Alzheimer's/ Parkinson's disease and strokes), vision disorders (e.g.		
CC	conjunctivitis, glaucoma, cataracts and retinitis pigmentosa),		
CC	reproductive disorders (e.g. infertility, uterine fibroids, ecologic		
CC	pregnancies and impotence) and smooth muscle disorders (e.g. angina,		
CC	anaphylactic shock, Kearns-Sayre syndrome and hypertension). It can also		
CC	be used for gene therapy.		
XX			
SQ	Sequence 4228 BP; 793 A; 1365 C; 1324 G; 746 T; 0 other;		

[illegible]

Qy	105	CCGCAGAGGGTTCCTACCTCGAGTGGTCCGCCGCCACCTCCAGTCCAGGAACCGCAACC	244
Db	1087	ctgtgagagccacatcctggaaacccctcggccctctgctgagccgclgctgcc	1146
Qy	245	CAGCCAAAAGGCTTCTCTTTCTGCTGCTCACCATCGTCTTCTGCGCAGATCTGATGGCTG	304
Db	1147	atgcccaggggacatctcaggct--tgcaagggtggccccctccgaacttcgacatgcct	1204
Qy	305	AAGAGGGTGTCCCGCGCCCTGCCTCCAGAGGACGGCCCTTAACGCCGCACTCCCTCGCGC	364
Db	1205	atgagcgtgycgggatctccgtgctccctgcaggagaggcctccgggggtccctggcag	1264
Qy	365	CCACCCCTGCTGCCCGTCTCCAGGCCCTTTAATCTGACTTCGGAGCCCTTCGACTAC	423
Db	1265	ccccgcctcgagcccccaactcaggagcctcgtagcagcggcggcgcclgtgaatac	1323
RESULT 14			
X53491/C			
ID	X53491 standard; DNA; 114955 BP.		
AC	X53491;		
XX	XX		
XX	05-JUL-1999 (first entry)		
XX	XX		
DE	Human adenosine A1 receptor antisense oligonucleotide fragment.		
XX	XX		
KW	Antisense oligonucleotide; multiple target; antisense treatment;		
KW	impaired respiration; inflammation; lung disease;		
KW	pulmonary vasoconstriction; inflammation; allergic rhinitis;		
KW	acute asthma; allergy; asthma; impeded respiration;		
KW	respiratory distress syndrome; pain; cystic fibrosis;		
KW	pulmonary hypertension; pulmonary vasoconstriction; emphysema;		
KW	chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;		
KW	colon cancer; breast cancer; lung cancer; pancreatic cancer;		
KW	hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;		
KW	prostate cancer;; ss.		
XX	XX		
OS	Synthetic.		
XX	XX		
PN	WO9913886-A1.		
XX	XX		
PD	25-MAR-1999.		
XX	XX		
XX	17-SEP-1998; 98WO-US19419.		
PF	XX		
PR	09-JUN-1998; 98US-0093972.		
PR	17-SEP-1997; 97US-0059160.		
XX	XX		
PA	(UYEC-) UNIV EAST CAROLINA.		
XX	XX		
PT	Nyce JW;		
XX	XX		
DR	WPI; 1999-229400/19.		
XX	XX		
PT	New antisense oligonucleotides used in treatment of, e.g. pulmonary		
PT	vasoconstriction		
XX	XX		
PS	Disclosure; Page 37; 120pp; English.		
XX	XX		
CC	The specification describes antisense oligonucleotides (X52869-X55271)		
CC	directed against at least 2 mRNAs selected from target genes, coding and		
CC	non-coding regions of RNAs corresponding to target genes, gene		
CC	initiation codons, genomic flanking regions, intron-exon borders, the		
CC	5'-end, the 3'-end and the juxta-section between coding and non-coding		
CC	regions and all segments of RNAs encoding proteins associated with one		
CC	or more diseases, conditions or mixtures. The antisense oligonucleotides		
CC	may be derived from sequences X55272-74. These multiple target		
CC	oligonucleotides (specifically X55180-271) can be used for the antisense		
CC	treatment of diseases and conditions. Typical diseases and conditions		
CC	are those associated with impaired respiration and inflammation,		
CC	including lung diseases, pulmonary vasoconstriction, inflammation,		





Search completed: January 30, 2001, 21:29:29  
Job time: 19255 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 23:11:29 ; Search time 64.84 Seconds  
(without alignments)  
155.458 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156

Perfect score: 442

Sequence: 1 VRROLVPEENPAKRLFL.....EPSDYALDLSTFLOQHPAAF 86

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_15:\*

- 1: sp-archaea:\*
- 2: sp-bacteria:\*
- 3: sp-fungi:\*
- 4: sp-human:\*
- 5: sp-invertebrate:\*
- 6: sp-mammal:\*
- 7: sp-mhc:\*
- 8: sp-organelle:\*
- 9: sp-phage:\*
- 10: sp-plant:\*
- 11: sp-rodent:\*
- 12: sp-virus:\*
- 13: sp-vertebrate:\*
- 14: sp-unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	442	100.0	193	4	075353
2	84.5	19.1	637	4	094878
3	81.5	18.4	102	5	09T2D3
4	81.5	18.4	452	5	09UAS5
5	81.5	18.4	468	5	09XY65
6	80.5	18.2	302	12	09QM69
7	78	17.6	4957	4	014687
8	78	17.6	5262	4	014686
9	77	17.4	1568	4	095785
10	77	17.4	2424	5	09VZ48
11	76	17.2	148	2	09JZ87
12	76	17.2	801	5	09W3K6
13	76	17.2	1677	5	000805
14	74	16.7	114	12	09VYX8
15	74	16.7	955	11	088287
16	74	16.7	1040	5	09VH04
17	74	16.7	1561	11	088286
18	73.5	16.6	925	12	091338
19	73	16.5	1029	12	09YMX0

20	72	16.3	659	5	09VQ23
21	72	16.3	852	4	09NP71
22	71	16.1	359	5	09TYW7
23	71	16.1	1882	3	09P898
24	70.5	16.0	244	12	069126
25	70.5	16.0	1900	4	015022
26	70.5	16.0	2605	4	09UK25
27	70.5	16.0	2715	4	09UMN6
28	70	15.8	148	2	09JUV2
29	70	15.8	1233	2	09R132
30	70	15.8	2971	4	09Y5L9
31	69.5	15.7	277	4	09NXN6
32	69.5	15.7	562	4	09NVJ4
33	69.5	15.7	745	4	09P2N8
34	69.5	15.7	757	6	09N015
35	69.5	15.7	831	4	09UMS6
36	69.5	15.7	1887	3	012533
37	69	15.6	418	5	018503
38	69	15.6	646	4	09ULX6
39	69	15.6	646	4	09UGM0
40	69	15.6	1454	5	010463
41	69	15.6	2157	11	09Z1R1
42	69	15.6	2187	11	P70670
43	68.5	15.5	701	4	09NXX8
44	68.5	15.5	734	4	09NSE6
45	68.5	15.5	821	4	09UII8

#### ALIGNMENTS

##### RESULT 1

1D 075353 PRELIMINARY; PRT; 193 AA.  
AC 075353;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-MAY-1999 (TrEMBLrel. 10; Last annotation update)  
DE ANTI-DEATH PROTEIN.  
GN IEX-1L.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-98369175; Pubmed-9703517;  
RA Wu M.X., Ao Z., Prasad K.V.S., Wu R., Schlossman S.F.;  
RT "IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival.";  
RL Science 281:998-1001(1998).  
DR EMBL; AF039067; AAC32558.1; -;  
DR EMBL; AF071596; AAC72344.1; -;  
SQ SEQUENCE 193 AA; 21028 MW; 7927D9D3FFBC7C57 CRC64;

Query Match 100.0%; Score 442; DB 4; Length 193;  
Best Local Similarity 100.0%; Pred. No. 8.7e-40;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRROLVPEENPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASLAPTVPSPVLE 60  
|||||  
Db 108 VRROLVPEENPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASLAPTVPSPVLE 167

QY 61 PFNLTSPSPDYALDLSTFLOQHPAAF 86  
|||||  
Db 168 PFNLTSPSPDYALDLSTFLOQHPAAF 193

##### RESULT 2

094878  
ID 094878 PRELIMINARY; PRT; 637 AA.  
AC 094878;

```

DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE KIAA0781 PROTEIN (FRAGMENT).
GN KIAA0781.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
KA Kotani H., Nomura N., Ohara O.,
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RL for large proteins in vitro."
RL DNA Res. 5:277-286(1998).
DR ENBL; AB018324; BAA34501.1; -.
DR INTERPRO; IPR000449; -.
DR INTERPRO; IPR002965; -.
DR PRINTS; PR01217; PRICHEXTENSN.
FT NON_TER
I
SQ SEQUENCE 637 AA; 70516 MW; DEBFB005BA9B9CDD CRC64;

Query Match 19.1%; Score 84.5; DB 4; Length 637;
Best Local Similarity 33.3%; Pred. No. 0.41;
Matches 30; Conservative 12; Mismatches 27; Indels 21; Gaps 5;

OY 7 VEEPNAKRLFL-----LTIIVFCOILMAEGVP-----APLPEDAPNAASLAP----- 52
DB 423 LOEHRIQOKRLFLQKOSLOAQFNQMTAESYPOPSQOLPLPROETPPPSQQAFFSLT 482

OY 53 TPVSPVLEPNTSPSDYALDLSTFLQOH 82
DB 483 QPLSPVLEP-----SSEQMQY-----SPFLSQY 505

RESULT 3
OYTD3
ID OYTD3 PRELIMINARY; PRT; 102 AA.
AC OYTD3
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE R160.5 PROTEIN.
GN R160.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Petoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans".
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;

```

```

RA Nelson J., Gattung S.;
RT "The sequence of C. elegans cosmid R160.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF099001; AAC68734.1; -.
SQ SEQUENCE 102 AA; 11426 MW; AFD366C3932EB26C CRC64;

```

```

Query Match 18.4%; Score 81.5; DB 5; Length 102;
Best Local Similarity 24.7%; Pred. No. 0.13;
Matches 19; Conservative 19; Mismatches 28; Indels 11; Gaps 3;

OY 11 NPAKRLFL-----LTIIVFCOILMAEGVPAPLPEDAPNAASLAPTPVSPVLEPFN-- 63
DB 7 NPAIOFAFVVILVCALAVCICMLIERADGFCPLPPEELRRIVNTIP-PRTPIAAPVNR 65

OY 64 ---LTSEPSDYALDLST 77
DB 66 QVVIKIDPTESDVTMTS 82

```

```

RESULT 4
OYUAS5
ID OYUAS5 PRELIMINARY; PRT; 452 AA.
AC OYUAS5
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE DEVELOPMENTAL-SPECIFIC PROTEIN CONZAB (FRAGMENT).
GN CONZAB.
OS Euplotes crassus.
OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotida; Euplotidae;
OC Euplotes.
OX NCBI_TaxID=5936;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ST9, ST11;
RX MEDLINE=97206572; PubMed=9172827;
RA Ling Z., Ghosh S., Jacobs M.E., Klobutcher L.A.;
RT "Conjugation-specific genes in the ciliate Euplotes crassus: gene
RT expression from the old macronucleus.";
RL J. Eukaryot. Microbiol. 44:1-11(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ST9, ST11;
RA Jacobs M.E., Ling Z., Klobutcher L.A.;
RT "conzab encodes a novel and abundant protein targeted to the
RT developing macronucleus in Euplotes crassus.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF116196; AAD29625.1; -.
FT NON_TER
I
SQ SEQUENCE 452 AA; 49763 MW; ASC13C49ECA998EA CRC64;

```

```

Query Match 18.4%; Score 81.5; DB 5; Length 452;
Best Local Similarity 34.5%; Pred. No. 0.61;
Matches 30; Conservative 10; Mismatches 30; Indels 17; Gaps 5;

OY 3 RQLPVEEP-----NPAKRLFLLLTTFVFCOILMAEGVPAPLPEDAPNAASLAPT-----p 54
DB 316 RQSPPLPDILENPAKEATSSLSKMI---QETAIESITPQPP-----VSSIQPSAPVHP 367

OY 55 VSPVLEPFN-LTSEPSDYALDLSTFLQ 80
DB 368 SAPSLEPSSTLASEPRDIIPDPSAALK 394

```

```

RESULT 5
OYXY65

```

ID O9XY65 PRELIMINARY: PRT: 468 AA.  
AC O9XY65;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE CONZAB PROTEIN.  
GN CONZAB.  
OS Euplotes crassus.  
OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotida; Euplotidae;  
OC Euplotes.  
OC Euplotes.  
OX NCBI\_TaxID=5936;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G1;  
RX MEDLINE=97206572; PubMed=9172827;  
RA Ling Z., Ghosh S., Jacobs M.E., Klobutcher L.A.;  
RT "Conjugation-specific genes in the ciliate Euplotes crassus: gene  
expression from the old macronucleus."  
RL J. Eukaryot. Microbiol. 44:1-11(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G1;  
RX MEDLINE=98391748; PubMed=9722644;  
RA Klobutcher L.A., Gyga S.E., Podoloff J.D., Vermeesch J.R.,  
RA Price C.M., Tebeau C.M., Jahn C.L.;  
RT "Conserved DNA sequences adjacent to chromosome fragmentation and  
telomere addition sites in Euplotes crassus."  
RL Nucleic Acids Res. 26:4230-4240(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G1;  
RA Ling Z., Klobutcher L.A.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF061334; AAD31724.1;  
SQ SEQUENCE 468 AA; 51457 MW; 70996125317C592A CRC64;

Query Match 18.4%; Score 81.5; DB 5; Length 468;  
Best Local Similarity 34.5%; Pred. No. 0.63;  
Matches 30; Conservative 10; Mismatches 30; Indels 17; Gaps 5;

OY 3 ROLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNNAASLAPT----P 54  
DB 332 RQSTPLDILENPAKEIASSLSKMI---QETAIESIPTQPP-----VSSIQPSAPVHP 383  
OY 55 VSPVLEPFN-LTSEPSDYALDLSTFLQ 80  
DB 384 SAPSLEPSTLASEPRDIIPDPSAALK 410

RESULT 6  
OY 09QM69 PRELIMINARY: PRT: 302 AA.  
AC 09QM69;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE 33 KDA PROTEIN.  
OS fowl adenovirus 8.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.  
OX NCBI\_TaxID=66295;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ojick D., Nagy E.;  
RT "The DNA sequence of fowl adenovirus 8."  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF083975; AAD50348.1;  
SQ SEQUENCE 302 AA; 34763 MW; BF34DF42FF026CCC CRC64;

Query Match 18.2%; Score 80.5; DB 12; Length 302;  
Best Local Similarity 26.4%; Pred. No. 0.51;  
Matches 24; Conservative 26; Mismatches 30; Indels 11; Gaps 4;

OY 2 RROLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLP-----PEDAPNNA-SLAPT 53  
DB 171 RKLLDLTDSDDSRSPFFMF-LVSDDAVIDEVPVPPPIPRKRRRRANKEEPNASTQIOLPE 229  
OY 54 PVSVPVLEPNLTSEPSDYALDLSTFLQHPA 84  
DB 230 PVSPAVS--DLKAEILNLLVEIESFVRKNPS 258

RESULT 7  
OY 014687 PRELIMINARY: PRT: 4957 AA.  
AC 014687;  
DT 01-JAN-1998 (TReMBLrel. 05, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
DE ALR.  
GN ALR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97388474; PubMed=9247308;  
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,  
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,  
RA Canaan E.;  
RT "Structure and expression pattern of human ALR, a novel gene with  
strong homology to ALL-1 involved in acute leukemia and to Drosophila  
trithorax."  
RL Oncogene 15:549-560(1997).  
DR EMBL; AF010404; AAC51735.1;  
DR INTERPRO; IPR001214;  
DR INTERPRO; IPR001822;  
DR INTERPRO; IPR001965;  
DR INTERPRO; IPR002965;  
DR PFAM; PF00628; PHD; 3;  
DR PFAM; PF00856; SET; 1;  
DR PRINTS; PR01217; PRICHEXTENS.  
DR PROSITE; PS00398; RECOMBINASES\_2; UNKNOWN1.  
SQ SEQUENCE 4957 AA; 531848 MW; 1026562E1419CE8D CRC64;

Query Match 17.6%; Score 78; DB 4; Length 4957;  
Best Local Similarity 30.4%; Pred. No. 17;  
Matches 21; Conservative 12; Mismatches 28; Indels 8; Gaps 2;

OY 1 VRROLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNNAASLAPTVPSPVLE 60  
DB 4299 LKQESPAPEPTQHRYTYNVSNLDVROL-----SAPPEEPSPPSPPLAPSAPSPTE 4351  
OY 61 PF-NLTSEP 68  
DB 4352 PLVELPTEP 4360

RESULT 8  
OY 014686 PRELIMINARY: PRT: 5262 AA.  
AC 014686;  
DT 01-JAN-1998 (TReMBLrel. 05, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
DE ALR.  
GN ALR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.

XX	MEDLINE=97388474; PubMed=9247308;
RA	Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA	Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA	Canaani E.;
RT	"Structure and expression pattern of human ALR, a novel gene with
RT	strong homology to ALL-1 involved in acute leukemia and to Drosophila
RT	trithorax".;
RL	Oncogene 15:549-560(1997).
DR	ENBL: AF010403; AAC51734.1; -.
DR	INTERPRO: IPR001214; -.
DR	INTERPRO: IPR001822; -.
DR	INTERPRO: IPR001841; -.
DR	INTERPRO: IPR001965; -.
DR	INTERPRO: IPR002965; -.
DR	PFAM: PF00628; PHD; 5.
DR	PFAM: PF00856; SET; 1.
DR	PRINTS: PR01217; PRICHEXTENS.
DR	PROSITE: PS00398; RECOMBINASE_2; UNKNOWN_1.
SQ	SEQUENCE 5262 AA; 564181 MW; 26B7C7CA0417E44 CRC64

Query Match 17.6%; Score 78; DB 4; Length 5262;  
Best Local Similarity 30.4%; Pred. No. 18;  
Matches 21; Conservative 13; Mismatches 28; Indels 8; Gaps 2;

Qy	1	VRRLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPILPPEDAPNAASLAPTVPSPVLE	60
		:::       : : :       :     :     :	
Db	4604	LKQESAPPEPTQHRYTYNVSNLDVRQL-----SAPPEEPSPPSP LAPSPASPTTE	4656

Qy 61 PF-NLTSEP 68

Db 4657 PLVELPTEP 4665

RESULT	9
O95785	
ID	O95785 PRELIMINARY; PRT; 1568 AA.
AC	
DT	01-MAY-1999 (TrEMBLrel. 10, Created)
DT	01-WAR-1999 (TrEMBLrel. 10, Last sequence update)
DT	01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE	HUMAN HOMOLOG OF MUS MUSCULUS WIZL PROTEIN (HUMAN HOMOLOG OF MUS
DE	MUSCULUS WIZS PROTEIN) (FRAGMENT).
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX	NCBI_TaxID=9606.

SEQUENCE FROM N.A.  
Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,  
Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,  
Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,  
Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,  
Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefficient J.,  
Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,  
Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,  
Olson A.S., Carrano A.V.:  
"Sequence analysis of a 3.5 Mb region in 19p13.1 between OLFR and  
JUND".  
Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

SEQUENCE OF 783-1568 FROM N.A.  
 Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,  
 Burkhardt-Schultz K., Gordon L., Dias J., Brower A., Ramirez M.,  
 Stilwagen S., Phan H., Velasco N., Do L., Regala W., Terry A.,  
 Carlens J., Danganan L., Erler A., Christensen M., Georgescu A.,  
 Avila J., Liu S., Attix C., Andreise T., Frankheim M.,  
 Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R.,  
 Thomas P., Quan G., Konniller B., Arellano A., Sanders C., Ow D.,  
 Nolan M., Trong S., Kobayashi A., Olsen A.S., Carraro A.V.;  
 "Sequence analysis of an ~1.5 Mb contig in 19p13.1 between OLF1 and  
 D19S885".  
 Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

```

DR EMBL; AC007059; AAd19818.1; -
DR EMBL; AC006128; AAC97985.1; -
DR EMBL; AC007059; AAd19817.1; -
DR INTERPRO; IPR000822; -
DR PFAM; PF00096; zf-C2H2; 10.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 8.
KW zinc-finger; Metal-binding; DNA-binding.
FT NON_TER 1
SQ SEQUENCE 1568 AA; 170077 MW; 74F0D34DA5F565D0 CRC64;

Query Match 17.4%; Score 77; DB 4; length 1568;
Best Local Similarity 40.4%; Pred. No. 6.6;
Matches 19; Conservative 3; Mismatches 17; Indels 8; Gaps 1

Oy 30 MAEGVPAIPUPPEDAPNAASLAPTPVSPVLE-----PFNLTSEP 68
: : : : : : : : : : : : : : : :
Db 908 LALAGSTPKNPEDKSPQLSILRPSAPSKAQPOSDEGPLNLTSGP 954
: : : : : : : : : : : : : : : :

```

RESULT	10	
Q9VZ48		
ID	Q9VZ48	PRELIMINARY; PRT; 2424 AA.
AC	Q9VZ48;	
DT	01-MAY-2000	(TREMBlrel. 13, Created)
DT	01-MAY-2000	(TREMBlrel. 13, Last sequence update)
DT	01-OCT-2000	(TREMBlrel. 15, Last annotation update)
DE	CG2174	PROTEIN.
GN	CG2174.	
OS	Drosophila melanogaster	(Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	

[1]  
SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananatlides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Abgaryan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.G.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bereman D.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Boudreau D., Bolshakov S.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieux P., Brotter P.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Davis P.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W.,  
RA Foster C., Gabriellista A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Hostin N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., MCPerson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeter F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong K., Sun E.,  
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.F., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003484; AAF47980.1; -.
DR HSP: P08799; IMND.
DR FLYBASE: FBgn0030252; CG2174.
DR INTERPRO: IPR000048; -.
DR INTERPRO: IPR000857; -.
DR INTERPRO: IPR001609; -.
DR PFAM: PF00063; myosin_head: 1.
DR PFAM: PF00612; 10; 3.
DR PFAM: PF00784; MYTH4; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
SQ SEQUENCE 2424 AA; 267616 MW; 8AD62AA33F9AA5D0 CRC64;

Query Match
Best Local Similarity 17.4%; Score 77; DB 5; Length 2424;
Matches 22; Conservative 7; Mismatches 29; Indels 2; Gaps 2;

QY 28 ILMAEGVPAPLPEDAPNAASLAPTPVSPVLEPENLTSEPS-DVALDLSTFLOOHPAAF 86
DB 1950 LLPPAPVPAPPPPIRPSMAPPAPPP-APQSPPTARSPPEPDITRTSSQVVKVCHVPAPF 2008

RESULT 11
Q9JZK7
ID Q9JZK7 PRELIMINARY; PRT: 148 AA.
AC Q9JZK7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ACYL COA THIOESTER HYDROLASE FAMILY PROTEIN.
GN NMB0925.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=2017555; PubMed=10710307;
RA Tottelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathavan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58.";
RL Science 287:1809-1815(2000).
DR EMBL: AE002444; AAF41333.1; -.
DR TIGR: NMB0925; -.
KW Hydrolase.
SQ SEQUENCE 148 AA; 16332 MW; 28E9B40D502D80B4 CRC64;

Query Match
Best Local Similarity 17.2%; Score 76; DB 2; Length 148;
Matches 15; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY 18 FLLITIVFCQILMAEGVPAPLPEDAPNAASLAPTP 54
DB 112 VLVTEAVTYVAIDAGNPRPIPEGNPKLAGLLPTP 148

RESULT 12
Q9W3K6
ID Q9W3K6 PRELIMINARY; PRT: 801 AA.
AC Q9W3K6
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

```

```

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DR CG2258 PROTEIN.
GN CG2258.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,
RA Ballev R.M., Basu A., Raxendale J., Bayraktaroglu L., Heasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Hochs S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,
RA Shue B.C., Siden-Klamos I., Simpson M.D., Skupski M.P., Shen T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003442; AAF46320.1; -.
DR HSP: P06241; 15HF.
DR FLYBASE: FBgn002997; CG2258.
DR INTERPRO: IPR001452; -.
DR PFAM: PF00018; SH3; 1.
DR PROSITE: PS50002; SH3; 1.
SQ SEQUENCE 801 AA; 89233 MW; A0F5C824663DC391 CRC64;

Query Match
Best Local Similarity 17.2%; Score 76; DB 5; Length 801;
Matches 21; Conservative 8; Mismatches 23; Indels 10; Gaps 2;

QY 26 CQILMAEGVPAPLPEDAPNAASLAPTPVS-----PVLEPNLTSEPSDYALDL 75
DB 206 CTIOTPOQQQLPAPPAPAPSPSAASATPTAASTKAATCGDVAQDNIITSEPSFTTTT 265

RESULT 13
QY 76 ST 77
DB 266 TT 267

```

```
000805
ID O00805 PRELIMINARY: PRT: 1677 AA.
AC O00805
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE P-GLYCOPROTEIN E.
OS Leishmania tropica.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LRC-L39;
RA Lafuente E., Castanys S., Gamarro F.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U55381; AA851191.1;
DR HSSP: P13569; INBD.
DR INTERPRO: IPR000504;
DR INTERPRO: IPR001140;
DR INTERPRO: IPR001167;
DR INTERPRO: IPR001865;
DR PFAM: PF00005; ABC_tran; 2.
DR PFAM: PF00664; ABC_membrane; 2.
DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE: PS00962; RIBOSOMAL_S2_1; UNKNOWN_1.
SQ SEQUENCE 1677 AA; 183006 MW; C49208921871563A CRC64;

Query Match 17.2%; Score 76; DB 5; Length 1677;
Best Local Similarity 33.9%; Pred. No. 9;
Matches 20; Conservative 6; Mismatches 15; Indels 18; Gaps 2;

QY 37 APLP-----PEDAPNAASLAPTVSPVLEPFN-----LTSEPSDYALDLST 77
DB 698 APLPEAEFLDPQRPSSSAAPRAGVPVTEPLNAKGSHTGHDAAASSEPLSSAQKST 756

RESULT 14
QYVYQ8
ID Q9VYQ8 PRELIMINARY: PRT: 114 AA.
AC Q9VYQ8
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE LATE 33 K PROTEIN HOMOLOG.
OS fowl adenovirus 8.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.
OX NCBI_TaxID=66295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC A-2A;
RX MEDLINE=98451338; PubMed=9780058;
RA Cao J.X., Krell P.J., Nagy E.;
RT "Sequence and transcriptional analysis of terminal regions of the fowl
RL adenovirus type 8 genome."
RL J. Gen. Virol. 79:2507-2516(1998).
DR EMBL: AF021254; AAC71673.1;
SQ SEQUENCE 114 AA; 13266 MW; 866DEEC6636EABA6 CRC64;

Query Match 16.7%; Score 74; DB 12; Length 114;
Best Local Similarity 31.1%; Pred. No. 0.94;
Matches 19; Conservative 17; Mismatches 15; Indels 10; Gaps 3;

QY 32 EEGVPAPLP-----PEDAPNAA-SLAPTVSPVLEPFNLTSEPSDYALDLSTFLOQHP 83
DB 12 EEPVPPPLPKRRRRRANKKEPNASSETOLPEVPSPAVS--DLKAEILNLLVEIESFVRKNP 69

QY 84 A 84
DB 70 S 70
```

```
RESULT 15
O88287
ID O88287 PRELIMINARY: PRT: 955 AA.
AC O88287
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE WIZ.
GN WIZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Matsumoto K., Ishii N., Yoshida S., Shiosaka S., Wanaka A.,
RA Tohyama M.;
RT "Molecular Cloning and Distinct Developmental Expression Pattern of
RL Spliced Forms of A Novel Zinc Finger Gene wiz in The Cerebellum.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB012266; BAA32791.1;
DR MGD: MGI:1332638; WIZ.
DR INTERPRO: IPR000822;
DR PFAM: PF00096; zf-C2H2; 5.
DR PRINTS: PR00048; ZINC_FINGER.
DR PROSITE: PS00028; ZINC_FINGER_C2H2; 5.
KW Zinc-finger; Metal-binding; DNA-binding.
SQ SEQUENCE 955 AA; 102777 MW; D4DD03A02BECCAAC CRC64;

Query Match 16.7%; Score 74; DB 11; Length 955;
Best Local Similarity 41.9%; Pred. No. 8.3;
Matches 18; Conservative 2; Mismatches 15; Indels 8; Gaps 1;

QY 34 GVPAPLPEDAPNNAASLAPTVSPVLE-----PFNLTSEP 68
DB 300 GSPTPKNPEDKSPOLSLSPRTSPRAQCQSEDEGPLNLTSGP 342
```

Search completed: January 30, 2001, 23:11:30  
Job time: 3995 sec



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 21:29:34 ; Search time 57.93 seconds  
(without alignments)  
401.549 Million cell updates/sec

Title: US-08-799-910-10

Perfect score: 823

Sequence: 1 MCHSRCHPTMTILOAPTPA.....EPSYALDSTFLOQHPAAF 156

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 912318 seqs, 149113690 residues

Total number of hits satisfying chosen parameters: 912318

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pcp.\*
- 2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pcp.\*
- 3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pcp.\*
- 4: /cgn2\_6/ptodata/2/paa/US080\_COMB.pcp.\*
- 5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pcp.\*
- 6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pcp.\*
- 7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pcp.\*
- 8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pcp.\*
- 9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pcp.\*
- 10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pcp.\*
- 11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pcp.\*
- 12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pcp.\*
- 13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pcp.\*
- 14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pcp.\*
- 15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pcp.\*
- 16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pcp.\*
- 17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pcp.\*
- 18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pcp.\*
- 19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pcp.\*
- 20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pcp.\*
- 21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pcp.\*
- 22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pcp.\*
- 23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pcp.\*
- 24: /cgn2\_6/ptodata/2/paa/US100\_COMB.pcp.\*
- 25: /cgn2\_6/ptodata/2/paa/US101\_COMB.pcp.\*
- 26: /cgn2\_6/ptodata/2/paa/US102\_COMB.pcp.\*
- 27: /cgn2\_6/ptodata/2/paa/US103\_COMB.pcp.\*
- 28: /cgn2\_6/ptodata/2/paa/US104\_COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	823	100.0	156	11	US-08-799-910-10
2	823	100.0	156	12	US-08-825-486-10
3	823	100.0	156	12	US-08-826-248-10
4	823	100.0	156	12	US-08-870-434-11
5	823	100.0	156	13	US-08-925-588-10

6	823	100.0	156	17	US-09-372-044-10
7	794	96.5	151	22	US-60-197-873-17386
8	440	53.5	123	22	US-60-196-718-6102
9	425	51.6	83	1	PCT-US00-15136-168
10	423	51.4	83	1	PCT-US00-15136-96
11	422	51.3	83	1	PCT-US00-15136-167
12	381	46.3	89	22	US-60-196-718-4945
13	358	43.5	71	1	PCT-US00-15136-169
14	358	43.5	92	22	US-60-196-718-5439
15	340	41.3	72	1	PCT-US00-15136-170
16	273.5	33.2	56	22	US-60-196-718-3936
17	114.5	13.9	1828	22	US-60-188-986-15
18	114.5	13.9	2012	22	US-60-229-524-80
19	114.5	13.9	2972	19	US-09-579-181-2
20	114.5	13.9	3041	28	US-60-258-275-409
21	114.5	13.9	3118	19	US-09-579-181-1
22	112.5	13.7	649	22	US-60-191-637-28141
23	112.5	13.7	649	22	US-60-191-681-22750
24	112.5	13.7	1255	22	US-60-167-217-13778
25	112.5	13.7	1255	22	US-60-173-464-11201
26	112	13.6	747	18	US-09-417-507-31296
27	110	13.4	590	1	PCT-US99-228538-2801
28	109	13.2	8991	11	US-08-714-741-32
29	108	13.1	271	16	US-09-248-796-21050
30	108	13.1	271	22	US-60-096-409-21050
31	107.5	13.1	1013	22	US-60-191-637-39541
32	107.5	13.1	1013	22	US-60-191-681-30574
33	107	13.0	206	9	US-08-529-055-54
34	106	12.9	1110	22	US-60-243-468-898
35	105.5	12.8	1013	22	US-60-173-464-30071
36	105	12.8	296	22	US-60-173-464-18776
37	105	12.8	296	22	US-60-191-637-22907
38	105	12.8	296	22	US-60-191-681-18056
39	104	12.6	129	9	US-08-529-055-53
40	102	12.4	331	22	US-60-173-464-29642
41	102	12.4	331	22	US-60-191-637-38878
42	102	12.4	331	22	US-60-191-681-30133
43	102	12.4	421	18	US-09-451-320-4244
44	101.5	12.3	193	9	US-08-529-055-49
45	101.5	12.3	741	18	US-09-417-507-37506

## ALIGNMENTS

RESULT 1  
US-08-799-910-10  
; Sequence 10, Application US/08799910  
; GENERAL INFORMATION:  
; APPLICANT: Falb, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/799,910  
; FILING DATE: 13-FEB-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/011,787  
; FILING DATE: 16-FEB-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-067-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 156 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-799-910-10

Query Match 100.0%; Score 823; DB 11; Length 156;  
Best Local Similarity 100.0%; Pred. No. 3.2e-60;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MCHSRCHPTMTILQAPTPAPSTIPGPRRSGPGEITFDPLPEPAAAPAGRPASRGRHK 60  
DB 1 MCHSRCHPTMTILQAPTPAPSTIPGPRRSGPGEITFDPLPEPAAAPAGRPASRGRHK 60  
QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120  
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120  
QY 121 APTVPSPVLEPFLNTSEPSDYALDSTFLOQHPAAF 156  
DB 121 APTVPSPVLEPFLNTSEPSDYALDSTFLOQHPAAF 156

RESULT 2  
US-08-825-486-10  
Sequence 10, Application US/08825486  
GENERAL INFORMATION:  
APPLICANT: Falb, Dean  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
THE TREATMENT AND DIAGNOSIS OF  
TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/825,486  
FILING DATE: 28-MAR-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/799,910  
FILING DATE: 13-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-077-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:  
LENGTH: 156 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-825-486-10  
Query Match 100.0%; Score 823; DB 12; Length 156;  
Best Local Similarity 100.0%; Pred. No. 3.2e-60;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MCHSRCHPTMTILQAPTPAPSTIPGPRRSGPGEITFDPLPEPAAAPAGRPASRGRHK 60  
DB 1 MCHSRCHPTMTILQAPTPAPSTIPGPRRSGPGEITFDPLPEPAAAPAGRPASRGRHK 60  
QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120  
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120  
QY 121 APTVPSPVLEPFLNTSEPSDYALDSTFLOQHPAAF 156  
DB 121 APTVPSPVLEPFLNTSEPSDYALDSTFLOQHPAAF 156

RESULT 3  
US-08-826-248-10  
Sequence 10, Application US/08826248  
GENERAL INFORMATION:  
APPLICANT: Falb, Dean  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
THE TREATMENT AND DIAGNOSIS OF  
TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/826,248  
FILING DATE: 28-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/799,910  
FILING DATE: 13-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-079-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 156 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein

```
; FRAGMENT TYPE: internal
US-08-826-248-10

Query Match 100.0%; Score 823; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIITFDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIITFDPLPEPAAAPAGRPSASRGHRK 60

Qy 61 RSRRLYPRVVRROLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120
Db 61 RSRRLYPRVVRROLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120

Qy 121 APTVPSPVLEPFLNTSEPSDYALDLSLTFLOQHPPAAF 156
Db 121 APTVPSPVLEPFLNTSEPSDYALDLSLTFLOQHPPAAF 156

RESULT 5
US-08-925-588-10
; Sequence 10, Application US/08925588
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; THE TREATMENT AND DIAGNOSIS OF
; CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,588
; FILING DATE: 08-Sep-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-08-925-588-10

Query Match 100.0%; Score 823; DB 13; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIITFDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIITFDPLPEPAAAPAGRPSASRGHRK 60

Qy 61 RSRRLYPRVVRROLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120
Db 61 RSRRLYPRVVRROLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120

Qy 121 APTVPSPVLEPFLNTSEPSDYALDLSLTFLOQHPPAAF 156
Db 121 APTVPSPVLEPFLNTSEPSDYALDLSLTFLOQHPPAAF 156

US-08-870-434-11
; Sequence 11, Application US/08870434
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,434
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-084
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-870-434-11

Query Match 100.0%; Score 823; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIITFDPLPEPAAAPAGRPSASRGHRK 60
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIITFDPLPEPAAAPAGRPSASRGHRK 60
```

```
RESULT 6
US-09-372-044-10
: Sequence 10, Application US/09372044A
: GENERAL INFORMATION:
: APPLICANT: Dean FALB et al.
: TITLE OF INVENTION: Compositions and Methods for the
: TITLE OF INVENTION: Treatment and Diagnosis of Cardiovascular Disease
: FILE REFERENCE: 7853-152
: CURRENT APPLICATION NUMBER: US/09/372,044A
: CURRENT FILING DATE: 1999-08-11
: NUMBER OF SEQ ID NOS: 44
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 10
: LENGTH: 156
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-372-044-10

Query Match      100.08; Score 823; DB 17; Length 156;
Best Local Similarity 100.08; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILQAPTPAPSTIPGRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILQAPTPAPSTIPGRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVRRQLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASL 120
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASL 120

QY 121 APTVPSPVLEPNTSEPSDYALDLSTFLQHPAAF 156
DB 121 APTVPSPVLEPNTSEPSDYALDLSTFLQHPAAF 156

RESULT 7
US-60-197-873-17386
: Sequence 17386, Application US/60197873
: GENERAL INFORMATION:
: APPLICANT: Beljano, Stephanie
: APPLICANT: Tanaka, Hiroaki
: APPLICANT: Dumas Milne Edwards, Jean Baptiste
: APPLICANT: Jobert, Severin
: APPLICANT: Giordano, Jean-Yves
: TITLE OF INVENTION: ESTs and Encoded Human Proteins.
: CURRENT APPLICATION NUMBER: US/60/197,873
: CURRENT FILING DATE: 2000-04-18
: NUMBER OF SEQ ID NOS: 52153
: SOFTWARE: Patent.pm
: SEQ ID NO 17386
: LENGTH: 151
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-197-873-17386

Query Match      96.58; Score 794; DB 22; Length 151;
Best Local Similarity 100.0%; Pred. No. 7.5e-58;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILQAPTPAPSTIPGRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILQAPTPAPSTIPGRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVRRQLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASL 120
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASL 120

QY 121 APTVPSPVLEPNTSEPSDYALDLSTFLQ 151
```

```
DB 121 APTVPSPVLEPNTSEPSDYALDLSTFLQ 151

RESULT 8
US-60-196-718-6102
: Sequence 6102, Application US/60196718
: GENERAL INFORMATION:
: APPLICANT: Bonazzi, Vivien
: TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AN
: TITLE OF INVENTION: USES THEREOF
: FILE REFERENCE: CL000456
: CURRENT APPLICATION NUMBER: US/60/196,718
: CURRENT FILING DATE: 2000-04-13
: NUMBER OF SEQ ID NOS: 7494
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 6102
: LENGTH: 123
: TYPE: PRT
: ORGANISM: HUMAN
US-60-196-718-6102

Query Match      53.58; Score 440; DB 22; Length 123;
Best Local Similarity 93.58; Pred. No. 7.4e-29;
Matches 86; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 65 VLYPRVRRQLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASLPTP 124
DB 32 LFLCPLRRQLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASLPTP 91

QY 125 VSPVLEPNTSEPSDYALDLSTFLQHPAAF 156
DB 92 VSPVLEPNTSEPSDYALDLSTFLQHPAAF 123

RESULT 9
PCT-US00-15136-168
: Sequence 168, Application PC/TUS0015136
: GENERAL INFORMATION:
: APPLICANT: Human Genome Sciences, Inc.
: TITLE OF INVENTION: 50 Human Secreted Proteins
: FILE REFERENCE: PS551PCT
: CURRENT APPLICATION NUMBER: PCT/US00/15136
: CURRENT FILING DATE: 2000-06-01
: EARLIER APPLICATION NUMBER: 60/138,629
: EARLIER FILING DATE: 1999-06-11
: NUMBER OF SEQ ID NOS: 198
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 168
: LENGTH: 83
: TYPE: PRT
: ORGANISM: Homo sapiens
PCT-US00-15136-168

Query Match      51.68; Score 425; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 8.2e-28;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 VVRRLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASLPTPVSPL 129
DB 1 VVRRLPVEEPNPAKRLLELLTIVFCQILMAEEGVPAPLPPEDAPNAASLPTPVSPL 129

QY 130 EPFNTSEPSDYALDLSTFLQ 152
DB 61 EPFNTSEPSDYALDLSTFLQ 83

RESULT 10
PCT-US00-15136-96
: Sequence 96, Application PC/TUS0015136
```

; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: 50 Human Secreted Proteins  
; FILE REFERENCE: PSS51PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/15136  
; CURRENT FILING DATE: 2000-06-01  
; EARLIER APPLICATION NUMBER: 60/138,629  
; EARLIER FILING DATE: 1999-06-11  
; NUMBER OF SEQ ID NOS: 198  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 96  
; LENGTH: 83  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (83)  
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
PCT-US00-15136-96

Query Match 51.4%; Score 423; DB 1; Length 83;  
Best Local Similarity 100.0%; Pred. No. 1.2e-27;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 75 LPVEEPNPAKRLFLLLTIVFCOILMAEEGVPAIPPPEDAPNAASLAPTPVSPVLEPRL 134  
Db 1 LPVEEPNPAKRLFLLLTIVFCOILMAEEGVPAIPPPEDAPNAASLAPTPVSPVLEPRL 60  
Qy 135 TSEPSDYALDLSTFLQHPAAF 156  
Db 61 TSEPSDYALDLSTFLQHPAAF 82

RESULT 11  
PCT-US00-15136-167  
; Sequence 167, Application PC/TUS0015136  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: 50 Human Secreted Proteins  
; FILE REFERENCE: PSS51PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/15136  
; CURRENT FILING DATE: 2000-06-01  
; EARLIER APPLICATION NUMBER: 60/138,629  
; EARLIER FILING DATE: 1999-06-11  
; NUMBER OF SEQ ID NOS: 198  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 167  
; LENGTH: 83  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US00-15136-167

Query Match 51.3%; Score 422; DB 1; Length 83;  
Best Local Similarity 98.8%; Pred. No. 1.4e-27;  
Matches 82; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 70 VYRROLPVVEEPNPAKRLFLLLTIVFCOILMAEEGVPAIPPPEDAPNAASLAPTPVSPVL 129  
Db 1 LVYRROLPVVEEPNPAKRLFLLLTIVFCOILMAEEGVPAIPPPEDAPNAASLAPTPVSPVL 60  
Qy 130 EPNLTSEPSDYALDLSTFLQOH 152  
Db 61 EPNLTSEPSDYALDLSTFLQOH 83

RESULT 12  
US-60-196-718-4945  
; Sequence 4945, Application US/60196718  
; GENERAL INFORMATION:  
; APPLICANT: Bonazzi, Vivien  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AT  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL000456  
; CURRENT APPLICATION NUMBER: US/60/196,718  
; CURRENT FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7494  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 4945  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: HUMAN  
; US-60-196-718-4945

Query Match 46.3%; Score 381; DB 22; Length 89;  
Best Local Similarity 100.0%; Pred. No. 3.7e-24;  
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRSCHPTMTILOAPTAPSTIPGPRGSGPEITFDPLPEPAAAPAGPSASRGHRK 60  
Db 1 MCHSRSCHPTMTILOAPTAPSTIPGPRGSGPEITFDPLPEPAAAPAGPSASRGHRK 60  
Qy 61 RSRRLVLYPRV 70  
Db 61 RSRRLVLYPRV 70

RESULT 13  
PCT-US00-15136-169  
; Sequence 169, Application PC/TUS0015136  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: 50 Human Secreted Proteins  
; FILE REFERENCE: PSS51PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/15136  
; CURRENT FILING DATE: 2000-06-01  
; EARLIER APPLICATION NUMBER: 60/138,629  
; EARLIER FILING DATE: 1999-06-11  
; NUMBER OF SEQ ID NOS: 198  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 169  
; LENGTH: 71  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US00-15136-169

Query Match 43.5%; Score 358; DB 1; Length 71;  
Best Local Similarity 97.1%; Pred. No. 2.2e-22;  
Matches 66; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MCHSRSCHPTMTILOAPTAPSTIPGPRGSGPEITFDPLPEPAAAPAGPSASRGHRK 60  
Db 1 MCHSRSCHPTMTILOAPTAPSTIPGPRGSGPEITFDPLPEPAAAPAGPSASRGHRK 60  
Qy 61 RSRRLVLYP 68  
Db 61 RSRRLVLYP 68

RESULT 14  
US-60-196-718-5439  
; Sequence 5439, Application US/60196718  
; GENERAL INFORMATION:  
; APPLICANT: Bonazzi, Vivien  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AT  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL000456  
; CURRENT APPLICATION NUMBER: US/60/196,718  
; CURRENT FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7494  
; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 5439  
; LENGTH: 92  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-60-196-718-5439

Query Match 43.5%; Score 158; DB 22; Length 92;  
Best Local Similarity 100.0%; Pred. No. 2.9e-22;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILOAPTAPSTIPGPRRGSGPEIFTDPLPEPAAAPAGRPASRGHRK 60  
|||||  
DB 1 MCHSRCHPTMTILOAPTAPSTIPGPRRGSGPEIFTDPLPEPAAAPAGRPASRGHRK 60  
|||||

QY 61 RSRRLV 66  
|||||  
DB 61 RSRRLV 66  
|||||

RESULT 15  
PCT-US00-15136-170  
; Sequence 170, Application PC/TUS0015136  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: 50 Human Secreted Proteins  
; FILE REFERENCE: PS551PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/15136  
; CURRENT FILING DATE: 2000-06-01  
; EARLIER APPLICATION NUMBER: 60/138,629  
; EARLIER FILING DATE: 1999-06-11  
; NUMBER OF SEQ ID NOS: 198  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 170  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US00-15136-170

Query Match 41.3%; Score 340; DB 1; Length 72;  
Best Local Similarity 100.0%; Pred. No. 6.8e-21;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILOAPTAPSTIPGPRRGSGPEIFTDPLPEPAAAPAGRPASRGHRK 60  
|||||  
DB 1 MCHSRCHPTMTILOAPTAPSTIPGPRRGSGPEIFTDPLPEPAAAPAGRPASRGHRK 60  
|||||

QY 61 RS 62  
||  
DB 61 RS 62  
||

Search completed: January 30, 2001, 23:10:11  
Job time: 6037 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: January 30, 2001, 19:24:41 ; Search time 36.91 Seconds  
(without alignments)  
75.895 Million cell updates/sec

Title: US-08-799-910-10  
Perfect score: 823  
Sequence: 1 MCHRSRCHPTMTILOAPTPA.....EPSDYALDLSTFLOQHPAAF 156

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues  
Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	823	100.0	156	3	US-08-826-246-10
2	823	100.0	156	3	US-08-940-495-10
3	823	100.0	156	3	US-09-126-640-11
4	95.5	11.6	739	3	US-09-035-648-24
5	95.5	11.6	1596	3	US-09-356-952-3
6	93.5	11.4	905	2	US-08-574-959A-9
7	93.5	11.4	1135	2	US-08-574-959A-7
8	92	11.2	1400	1	US-08-080-255-7
9	92	11.2	1400	3	US-08-465-713-7
10	92	11.2	1400	4	PCT-US93-05857-7
11	90.5	11.0	214	1	US-08-217-327-4
12	88	10.7	468	2	US-08-390-000A-7
13	88	10.7	477	1	US-08-444-734A-4
14	88	10.7	477	1	US-08-087-772A-16
15	87	10.6	2414	1	US-08-227-536-2
16	87	10.6	2414	4	PCT-US95-04682-6
17	85.5	10.4	610	1	US-07-821-717B-6
18	85.5	10.4	610	1	US-08-119-262B-6
19	85.5	10.4	610	1	US-08-135-929A-11
20	85.5	10.4	610	1	US-08-234-265A-11
21	85.5	10.4	774	2	US-08-231-193A-42
22	85.5	10.4	774	2	US-08-486-273A-42
23	85.5	10.4	774	3	US-08-480-474-42
24	85.5	10.4	774	3	US-08-940-086A-42
25	85.5	10.4	1214	2	US-08-231-193A-54
26	85.5	10.4	1214	2	US-08-486-273A-54
27	85.5	10.4	1214	3	US-08-480-474-54
28	85.5	10.4	1214	3	US-08-940-086A-54

29	85.5	10.4	1219	2	US-08-231-193A-50	Sequence 50, Appl
30	85.5	10.4	1219	2	US-08-486-273A-50	Sequence 50, Appl
31	85.5	10.4	1219	3	US-08-480-474-50	Sequence 50, Appl
32	85.5	10.4	1219	3	US-08-940-086A-50	Sequence 50, Appl
33	85.5	10.4	1231	2	US-08-231-193A-48	Sequence 48, Appl
34	85.5	10.4	1231	2	US-08-486-273A-48	Sequence 48, Appl
35	85.5	10.4	1231	3	US-08-940-086A-48	Sequence 48, Appl
36	85.5	10.4	1231	3	US-08-480-474-48	Sequence 48, Appl
37	85.5	10.4	1231	3	US-08-231-193A-6	Sequence 6, Appl
38	85.5	10.4	1236	2	US-08-486-273A-6	Sequence 6, Appl
39	85.5	10.4	1236	3	US-08-480-474-6	Sequence 6, Appl
40	85.5	10.4	1236	3	US-08-940-086A-6	Sequence 6, Appl
41	85.5	10.4	1239	2	US-08-231-193A-52	Sequence 52, Appl
42	85.5	10.4	1239	2	US-08-486-273A-52	Sequence 52, Appl
43	85.5	10.4	1239	3	US-08-480-474-52	Sequence 52, Appl
44	85.5	10.4	1239	3	US-08-940-086A-52	Sequence 52, Appl
45	85.5	10.4	1244	2	US-08-231-193A-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1  
US-08-826-246-10  
; Sequence 10, Application US/08826246  
; Patent No. 6048709  
; GENERAL INFORMATION:  
; APPLICANT: Falb, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/826,246  
; FILING DATE: 28-MAR-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/799,910  
; FILING DATE: 13-FEB-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/011,787  
; FILING DATE: 16-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-078-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)7909090  
; TELEFAX: (212)8699741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 156 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-826-246-10



Query Match 100.0%; Score 823; DB 3; Length 156;  
Best Local Similarity 100.0%; Pred. No. 4e-75;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60  
DB 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVVRROLVPEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPPEDAPNAASL 120  
DB 63 RSRRLVPRVVRROLVPEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOQHPPAAF 156  
DB 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOQHPPAAF 156

RESULT 2  
US-08-944-495-10  
; Sequence 10, Application US/08944495  
; Patent No. 6087477  
; GENERAL INFORMATION:  
; APPLICANT: Falb, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944,495  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/799,910  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-067-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)7909090  
; TELEFAX: (212)8699741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 156 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-944-495-10

Query Match 100.0%; Score 823; DB 3; Length 156;  
Best Local Similarity 100.0%; Pred. No. 4e-75;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60  
DB 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVVRROLVPEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPPEDAPNAASL 120  
DB 61 RSRRLVPRVVRROLVPEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOQHPPAAF 156  
DB 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOQHPPAAF 156

RESULT 3  
US-09-126-640-11  
; Sequence 11, Application US/09126640A  
; Patent No. 6099823  
; GENERAL INFORMATION:  
; APPLICANT: Falb, Dean A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE;  
; FILE REFERENCE: 7853-126  
; CURRENT APPLICATION NUMBER: US/09/126,640A  
; CURRENT FILING DATE: 1998-07-30  
; EARLIER APPLICATION NUMBER: 08/870,434  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 08/799,910  
; EARLIER FILING DATE: 1997-02-13  
; EARLIER APPLICATION NUMBER: 60/011,787  
; EARLIER FILING DATE: 1996-02-16  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 156  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-126-640-11

Query Match 100.0%; Score 823; DB 3; Length 156;  
Best Local Similarity 100.0%; Pred. No. 4e-75;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60  
DB 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVVRROLVPEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPPEDAPNAASL 120  
DB 61 RSRRLVPRVVRROLVPEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOQHPPAAF 156  
DB 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOQHPPAAF 156

RESULT 4  
US-09-035-648-24  
; Sequence 24, Application US/09035648  
; Patent No. 6100031  
; GENERAL INFORMATION:  
; APPLICANT: Shyjan, Andrew W.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL  
; TITLE OF INVENTION: GROWTH AND PROLIFERATION  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/035.648  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/818.829  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Meiklejohn, Ph.D., Anita L.  
REGISTRATION NUMBER: 35.283  
REFERENCE/DOCKET NUMBER: 07334/003001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 739 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-09-035-648-24

Query Match 11.6%; Score 95.5; DB 3; Length 739;

Best Local Similarity 27.0%; Pred. No. 0.089;  
Matches 47; Conservative 14; Mismatches 60; Indels 53; Gaps 7;

QY 16 ATTPAPSTIPG-----RRSGPEITFDLPPEAAAPAGRPS-----ASRGHR---- 59  
DB 264 APPYSAVTPPDATSRGVSSDIAGPAQPPWPQPAKSPAFYDSSERIASRDERISVP 323  
QY 60 -KRS-----RRVLPVVRQLPVEENPAKRLFLLL----- 91  
DB 324 AKRTGILQAKRRSTTKPMFTTEPKVSPNP--ELISLLQNSEGRKCTGAGDGSPEEDY 381  
QY 92 -----TIVFCQILMAEEGVAPLPPEAPNAASLAP-TPVSPVLEPFLNTSEP 138  
DB 382 LSLGAEACNFMSSSAKQKTPPVAPKPAVKSSSQVTPVSPVNSGVVAPTOP 435

## RESULT 5

US-09-356-952-3  
Sequence 3, Application US/09356952  
Patent No. 6117663

## GENERAL INFORMATION:

APPLICANT: Horlack-Sjodin, Ann  
APPLICANT: Margarit, S. M.  
APPLICANT: Bor-Sogli, Dafna  
APPLICANT: Cole, Philip  
APPLICANT: Kuriyan, John  
TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE  
TITLE OF INVENTION: THEROF  
FILE REFERENCE: 600-1-228N  
CURRENT APPLICATION NUMBER: US/09/356.952  
CURRENT FILING DATE: 1999-07-19  
EARLIER APPLICATION NUMBER: 60/093.631  
EARLIER FILING DATE: 1998-07-21  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 3  
LENGTH: 1596  
TYPE: PPT  
ORGANISM: Drosophila melanogaster  
US-09-356-952-3

Query Match 11.6%; Score 95.5; DB 3; Length 1596;

Best Local Similarity 28.3%; Pred. No. 0.22;  
Matches 52; Conservative 19; Mismatches 68; Indels 45; Gaps 13;

QY 4 SRSCHPTWTILOAPTPAPSTIPGPRRG--SGPEITFDLPPEAAAPAGRPSASGHKKR 61  
DB 1354 TESCADMAQKQAP-DAP-TLP-PRGELSPP-----PIPRNLNISTGISVLRSHGKS 1404  
QY 62 SRRV-----LYPR-----VVRRLPVE-----ENPAKRLFLLLTIVFCQILMAEEGV 105  
DB 1405 KEFVGNSLLPNTSSIMIRNSAIEKRAAATSQPNOAAGPISTTLVTVSQAVATDEPL 1464  
QY 106 PAPLPPEAPNAASLAP--PVSPVLEP--NLTSEP-----SDYALDLSTLQO----H 152  
DB 1465 PLPISP-----AASSSTTSPITPAWSPMNPISHPVESTSSSYAHLRMHQOQQOQTH 1519  
QY 153 PAAF 156  
DB 1520 PAIY 1523

## RESULT 6

US-08-574-959A-9  
Sequence 9, Application US/08574959A  
Patent No. 5962224

## GENERAL INFORMATION:

APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi  
APPLICANT: and Jack L. Strominger  
TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES  
TITLE OF INVENTION: AND USES THEREFOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESS: LAHIVE & COCKFIELD  
STREET: 60 State Street, Suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/574.959A  
FILING DATE: 19-DEC-95  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36.207  
REFERENCE/DOCKET NUMBER: DFN-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 905 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-574-959A-9

## Query Match

11.4%; Score 93.5; DB 2; Length 905;  
Best Local Similarity 23.1%; Pred. No. 0.18;

Matches 45; Conservative 17; Mismatches 62; Indels 71; Gaps 8;

QY 8 HPTMTILO-----APTAPSTIPGPRRGSGPE-----ITFDLPPEAAAPAGRPSAS 55  
DB 400 HPRVPPLOPMGFTCTPAVPVLLRPHRPSGPHRSILRAPCOWAPCPOAACPCSGMPMS 459  
QY 56 RG-----HRKRSKR--VLYP----- 68  
DB 460 AGVPSEPTSTANLLGSLRSPVCPPLLLPGPENHRAGSNEDFILAPSGTPPTPTIPD 519  
QY 69 -----RVVRRLPVEENPAKRLFLLLTIVFCQILMAEEGVAPLPPEAPNAASLAP 123  
DB 520 ETFGGRVPRPFAVHYDKEEASDVETSLSDSDSVIVPEGLP-PLPP---PPPSGATPP 575

QY 124 PVSVPLEPNLTSEP 138  
Db 576 PIAPGPP---TASP 587  
RESULT 7  
US-08-574-959A-7  
; Sequence 7, Application US/08574959A  
; Patent No. 5962224  
; GENERAL INFORMATION:  
; APPLICANT: Jaekyoon Shin, Insil Jounq, Ratna K. Vadlamudi  
; APPLICANT: and Jack L. Strominger  
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES  
; TITLE OF INVENTION: AND USES THEREFOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, Suite 510  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/574.959A  
; FILING DATE: 19-DEC-95  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragoras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: DFN-008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)227-5941  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1135 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-574-959A-7  
Query Match 11.4%; Score 93.5; DB 2; Length 1135;  
Best Local Similarity 23.1%; Pred. No. 0.23;  
Matches 45; Conservative 17; Mismatches 62; Indels 71; Gaps 8;  
QY 8 HPTMTILQ-----APTAPSTIPGRRSGPE-----IFTFDLPPEAAAPAGRPSAS 55  
Db 630 HPRVPPLOPMGPTCTTPAPVPLLRHPSGPHRSILRAPCPQWAPCPQACPCPSAGPWS 689  
QY 56 RG-----HRKRSRR--VLYP----- 68  
Db 690 AGVPSEPTWTANLLGLLSRPSVCPRLPLPGPENHRAGSNEDDILAPSGTPPTIPPD 749  
QY 69 -----KVRROLVPEPNPAKRLLELLTIVFCQILMAEEGVAPILPPEDAPNAASLAPT 123  
Db 750 ETEGGRVPRPAFVHYDKAESDVEISLESDSDSVIVVEGLP-PLPP---PPPSGATPP 805  
QY 124 PVSVPLEPNLTSEP 138  
Db 806 PIAPGPP---TASP 817  
RESULT 8  
US-08-080-255-7  
; Sequence 7, Application US/08080255  
; Patent No. 548970  
; GENERAL INFORMATION:  
; APPLICANT: Rowley, Janet D.  
; APPLICANT: Diaz, Manuel O.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA

; APPLICANT: Rowley, Janet D.  
; APPLICANT: Diaz, Manuel O.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/080.255  
; FILING DATE: 19930617  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165  
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 320-7200  
; TELEFAX: (512) 474-7577  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1400 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-080-255-7  
Query Match 11.2%; Score 92; DB 1; Length 1400;  
Best Local Similarity 22.1%; Pred. No. 0.42;  
Matches 31; Conservative 17; Mismatches 40; Indels 52; Gaps 5;  
QY 15 QAPTAPSTIPGRRSGPEIFTDPLPE-----PAAAPAGRPSASGHRKKS 62  
Db 465 QKPTPSAREDPAPKSSS-----EPPRKPEEKESEGNVSAPGESKQATTPASRKSS 518  
QY 63 RRVLYPRV-----RRQLPVEPNPAKRLLELLTIVFCQILMAEEGVAPILPP 111  
Db 519 KVSQPALVIPQPTTGPPEKPKVKTTPSEKK-----KQPPPP 558  
QY 112 EDAPNAA---SLAPTPVSPV 128  
Db 559 ESGPEQSKKKVAPRPSIPV 578  
RESULT 9  
US-08-465-713-7  
; Sequence 7, Application US/08465713  
; Patent No. 6121419  
; GENERAL INFORMATION:  
; APPLICANT: Rowley, Janet D.  
; APPLICANT: Diaz, Manuel O.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA

```

: APPLICATION NUMBER: 07/900,689
: FILING DATE: 17/06/92
: ATTORNEY/AGENT INFORMATION:
: NAME: Parker, David L.
: REGISTRATION NUMBER: 32,165
: REFERENCE/DOCKET NUMBER: ARCD:(
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (512) 320-7200
: TELEFAX: (512) 474-7577
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1400 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
PCT-US93-05857-7

Query Match 11.2% Score:
Best Local Similarity 22.1% Predicted:
Matches 31; Conservative 17; Mismatches 14

QY 15 QAPTAPSTIPGRRGSGPEITFDPLPDLPEE
Db 465 KQTPSAREDPAPKSSS-----EPPK
QY 63 RRVLYPRVV-----RRQLPVVEE
Db 519 KQVSQALVLPQPTTGGPRKEVPKTTPTT
QY 112 EDAPNAA---SLAPTPSPV 128
Db 559 ESGSEQSKQKVAPRPSIPV 578

RESULT 11
US-08-217-327-4
: Sequence 4, Application US/08217327
: Patent No. 5474925
: GENERAL INFORMATION:
: APPLICANT: John, Maliyakal E
: APPLICANT: Barton, Kenneth A
: TITLE OF INVENTION: Immobilized P
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Quarles and Brady
: STREET: P.O. Box 2113
: CITY: Madison
: STATE: WI
: COUNTRY: USA
: ZIP: 53701-2113
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/217,338
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/812,272
: FILING DATE: 19-DEC-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Seay, Nicholas J
: REGISTRATION NUMBER: 27,386
: REFERENCE/DOCKET NUMBER: 1122990
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 608-251-5000
: TELEFAX: 608-251-9166
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:

```

LENGTH: 214 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-217-327-4

Query Match  
Best Local Similarity 27.5%; DB 1; Length 214;  
Matches 38; Conservative 8; Mismatches 59; Indels 33; Gaps 5;

QY 9 PTMTLQATPIAPSTIPGPRGSGREIFTDP-----LPEAAAPAGRTSASRCHKRS 62  
DB 54 PTATPPPVSTPTPTSSPPPVATSPVPSTPPSPPTATPPASPPTATPPASPPTATP 113  
QY 63 RRVLYPRVVRQLPVEENPAKRLFLLLTIVFCOILMAEGVPAPLPDAPNAASLAP 122  
DB 114 PPASPPATP---PPATPPA-----TPPPATPPAPPLASPVAT 149  
QY 123 TP-VSPVLEPENLTSEPS 139  
DB 150 VPAISPQVTP--LTSPPA 165

## RESULT 12

US-08-390-000A-7  
Sequence 7, Application US/08390000A  
Patent No. 5985583

GENERAL INFORMATION:  
APPLICANT: Sealion, Stuart C.  
TITLE OF INVENTION: Cloning and Expression of  
TITLE OF INVENTION: Gonadotropin-Releasing Hormone Receptor  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/390,000A  
FILING DATE: 17-FEB-1995  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 6923-052  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869-8864/9741

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 468 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-390-000A-7

Query Match  
Best Local Similarity 24.2%; DB 2; Length 468;  
Matches 50; Conservative 19; Mismatches 44; Indels 94; Gaps 12;

QY 17 TPATPSTIPGPRGSGREIFTDPLPEAAAPAGRPSAS--RGURKRSRRLVPRVVRQ 74  
DB 264 PSPSPSPVTPAPAPPCP-----PRPAAAAATAPLANGRKRPSRLV-----ALREQ 311

QY 75 -----LPVEENPAKRL-----LPLLT-----LVFCO- 97  
DB 312 KALKTLGIINGVFTLCLWLPFFLANVVKAFHRELVPDLVFFNNWLYANSAPNIIYKRS 371  
QY 98 --ILMAEEGV-----PAPLPEDAPNAAS-----LA 121  
DB 372 PDFRKAFOGLLCCARRARRRATHGDRPRASGLARPGP-PD--SIGAASDDDDDUUGV 428  
QY 122 PTPVSPVLEPF-----NLTSPPSYALD 144  
DB 429 ATPPARILLEWACNGCGAAADSSSLD 455

## RESULT 13

US-08-444-734A-4  
Sequence 4, Application US/08444734A  
Patent No. 5610282

GENERAL INFORMATION:  
APPLICANT: Sibley, David R.  
APPLICANT: Monsma, Frederick J.  
APPLICANT: Mahan, Lawrence C.  
APPLICANT: McVittie, Loris D.  
TITLE OF INVENTION: cDNA encoding the rat D1 dopamine  
TITLE OF INVENTION: receptor linked to adenylyl cyclase activation and  
TITLE OF INVENTION: expression of the receptor protein in plasmid-transfected  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson and Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/444,734A  
FILING DATE:  
CLASSIFICATION: 530

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/029,917  
FILING DATE: 03-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/548,714  
FILING DATE: 06-JUL-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E.  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: NIH065.001FW1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 477 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
US-08-444-734A-4

Query Match  
Best Local Similarity 24.2%; DB 1; Length 477;  
Matches 50; Conservative 19; Mismatches 44; Indels 94; Gaps 12;





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 18:47:31; Search time 54.97 Seconds  
(without alignments)  
97.039 Million cell updates/sec

Title: US-08-799-910-10  
Perfect score: 823  
Sequence: 1 MCHSRCHPTMTILOAFTP.....EPDYALDLSTLQHPAAF 156

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 268485 seqs, 34193795 residues  
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: A\_Geneseq\_36.\*  
1: /cgn2\_2/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /cgn2\_2/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /cgn2\_2/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /cgn2\_2/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /cgn2\_2/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /cgn2\_2/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /cgn2\_2/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /cgn2\_2/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /cgn2\_2/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /cgn2\_2/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /cgn2\_2/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /cgn2\_2/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /cgn2\_2/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /cgn2\_2/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /cgn2\_2/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /cgn2\_2/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /cgn2\_2/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /cgn2\_2/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /cgn2\_2/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /cgn2\_2/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /cgn2\_2/gcgdata/geneseq/geneseq/AA2000.DAT.\*

prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result, being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	823	100.0	156	21	Y45017 Protein encoded by
2	816	99.1	156	18	W36006 Human fchd605 gene
3	107	13.0	206	18	W14574 Streptococcus pneu
4	104	12.6	129	18	W14573 Streptococcus pneu
5	101.5	12.3	190	18	W14569 Streptococcus pneu
6	101	12.3	180	18	W14582 Streptococcus pneu
7	101	12.3	183	18	W14570 Streptococcus pneu
8	100	12.2	783	19	W37151 Mouse neural Mena+
9	100	12.2	787	19	W37152 Mouse neural Mena+
10	100	12.2	802	19	W37153 Mouse neural Mena+
11	98.5	12.0	168	21	Y68732 Amino acid sequenc
12	96.5	11.7	179	21	Y68731 Amino acid sequenc

13	95.5	11.6	739	18	W25790
14	95.5	11.6	1596	21	Y68821
15	94.5	11.5	520	16	R78185
16	94	11.4	1290	20	Y06809
17	94	11.4	1290	21	Y32216
18	94	11.4	1291	20	Y01687
19	93.5	11.4	185	18	W14566
20	93.5	11.4	905	18	W31186
21	93.5	11.4	1135	18	W31185
22	92.5	11.2	572	18	W31855
23	92.5	11.2	763	18	W31852
24	92	11.2	166	18	W14568
25	92	11.2	167	18	W14575
26	92	11.2	1400	15	R44514
27	92	11.2	1400	17	R92705
28	92	11.2	3910	14	R38470
29	92	11.2	3910	16	R66462
30	91	11.1	1012	20	Y17406
31	90.5	11.0	214	17	R86913
32	90	10.9	171	20	Y29034
33	90	10.9	171	20	Y29074
34	90	10.9	254	20	Y29061
35	89.5	10.9	666	19	W72911
36	89.5	10.9	666	20	Y21928
37	89	10.8	788	21	Y54466
38	89	10.8	807	21	Y54467
39	89	10.8	1291	20	Y01689
40	89	10.8	1291	20	Y06810
41	89	10.8	3969	15	R52971
42	88.5	10.8	187	18	W14579
43	88.5	10.8	188	18	W14580
44	88.5	10.8	688	21	Y57898
45	88.5	10.8	711	19	W80602

ALIGNMENTS

RESULT 1  
Y45017  
ID Y45017 standard; Protein; 156 AA.  
XX AC Y45017:  
XX AC  
DT 31-MAY-2000 (first entry)  
XX AC  
DE Protein encoded by fchd605 gene.  
XX AC  
KW fchd605 gene; human; cardiovascular disease; oncogenic disorder;  
KW diabetic retinopathy; fibroproliferative disorder; atherosclerosis;  
KW TGF-beta signalling pathway; TGF; Transforming growth factor;  
KW pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;  
KW vascularisation; cytostatic; antidiabetic; ophthalmological.  
OS Homo sapiens.  
XX  
XX WO200006206-A1.  
XX  
XX PD 10-FEB-2000.  
XX  
XX PF 30-JUL-1999; 99WO-US17394.  
XX  
XX PR 30-JUL-1998; 98US-0126640.  
XX  
XX PA (MILL-) MILLENNIUM PHARM INC.  
XX  
XX PI Falb DA;  
XX  
XX DR WPI; 2000-205414/18.  
XX DR N-PSDB; 250711.  
XX  
XX PT Identifying substances for ameliorating symptoms of fibroproliferative  
diseases or oncogenic related disorders -

Gene 036 product d  
Amino acid sequenc  
Protein sequence o  
Human per gene pro  
Human per gene pro  
Protein encoded by  
Streptococcus pneu  
Human p160 polytyp  
Human p160 polytyp  
Mycobacterium tube  
Mycobacterium tube  
Streptococcus pneu  
Streptococcus pneu  
MLL amino acid seq  
MLL cDNA clone 14-  
ALL-1 protein. Ho  
ALL-1 (acute lymph  
human atrophin-1 r  
Cotton fiber-speci  
T. gondii immunoge  
T. gondii immunoge  
T. gondii immunoge  
Mycobacterium tube  
Amino acid sequenc  
Amino acid sequenc  
Amino acid sequenc  
Protein encoded by  
Mouse per gene pro  
Product of the cDN  
Streptococcus pneu  
Streptococcus pneu  
Human transmembran  
Human transcriptio



XX

Example; Fig 5; 214pp; English.

XX

XX The patent discloses methods for the treatment and diagnosis of cardiovascular diseases by novel human genes which are differentially expressed in different cardiovascular disease states. Compositions which can modify TGF-beta signalling pathway are identified by screening. These are used therapeutically to treat fibroproliferative and oncogenic disorders, especially TGF (transforming growth factor)-beta related disorders, including diabetic retinopathy, arteriosclerosis, pancreatic cancer, angiogenesis, inflammation, fibrosis, tumour growth and vascularisation. The present sequence is the protein product of fchd605 gene which is up-regulated in monocytes treated with oxidised LDL (low density lipoprotein) can be used to design cardiovascular disease treatment strategies. Depending on whether the up-regulation has a pathogenic or protective effect treatment methods can be designed to increase or decrease the activity of the protein product of the gene.

XX Sequence 156 AA;

Query Match 100.0%; Score 823; DB 21; Length 156;  
Best Local Similarity 100.0%; Pred. No. 7.4e-69;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEITFDPLPEPAAPAGRPSASRGRK 60  
|||||  
DB 1 mchsrchptmtlilqaptapstipgprgsgpeitfdplpepaapagrpsasrgrk 60  
OY 61 RSRRLVYPRVVRQIPVEEPNPAKRLLFLLITIVFCQILMAEGVPAPLPEDAPNAASL 120  
DB 61 rsrriyprvrrqipveepnpakrllfllitvfcqilmaegvpaplpapedapnaasl 120  
OY 121 APTPVSPVLEPNLTSEPSDYALDLSTFLOQHPAAF 156  
DB 121 aptpvspvlepnltsepsdyaldlstflqhpaaaf 156

RESULT 2

W36006  
ID W36006 standard; Protein; 156 AA.  
AC W36006;

DT 03-MAR-1998 (first entry)

DE Human Fchd605 gene product.

KW Fchd605 gene; differential expression; monocyte; human;  
KW foam cell; cardiovascular disease; arteriosclerosis; ischaemia;  
KW reperfusion; hypertension; restenosis; arterial inflammation;  
KW therapy; diagnosis; drug screening; marker.

OS Homo sapiens.

PN WO9730065-A1.

PD 21-AUG-1997.

PF 14-FEB-1997; 97WO-US02291.

PR 13-FEB-1997; 97US-0799910.

PR 16-FEB-1996; 96US-0011787.

PA (MILL-) MILLENNIUM PHARM INC.

PI Fatb DA;

DR WPI; 1997-424966/39.

XX N-PSDB; T94471.

XX New genes differentially expressed in cardiovascular disease - used  
PT for diagnosis, drug screening and treatment of cardiovascular

PT disease, e.g. atherosclerosis, restenosis, hypertension, etc

XX Example 6; Fig 5; 163pp; English.

XX This protein is encoded by the novel human fchd605 gene (see T94471) that is up-regulated in monocytes treated with oxidised low density lipoproteins that simulate the conditions under which foam cells develop during atherogenesis. The protein has sequence similarity to the mouse gly96 gene and to EST T49532. Novel fchd531, fchd540, fchd545, fchd602 and fchd605 genes (see T94467-71) provide a fingerprint for the study of cardiovascular diseases, including atherosclerosis, ischaemia/reperfusion, hypertension, restenosis and arterial inflammation. Methods are provided for the diagnosis, monitoring in clinical trials, screening for therapeutically effective compounds, and treatment of cardiovascular diseases based on discoveries regarding the expression patterns of these novel genes.

XX Sequence 156 AA;

Query Match 99.1%; Score 816; DB 18; Length 156;  
Best Local Similarity 99.4%; Pred. No. 3.3e-68;  
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEITFDPLPEPAAPAGRPSASRGRK 60  
|||||  
DB 1 mchsrchptmtlilqaptapstipgprgsgpeitfdplpepaapagrpsasrgrk 60  
OY 61 RSRRLVYPRVVRQIPVEEPNPAKRLLFLLITIVFCQILMAEGVPAPLPEDAPNAASL 120  
DB 61 rsrriyprvrrqipveepnpakrllfllitvfcqilmaegvpaplpapedapnaasl 120  
OY 121 APTPVSPVLEPNLTSEPSDYALDLSTFLOQHPAAF 156  
DB 121 aptpvspvlepnltsepsdyaldlstflqhpaaaf 156

RESULT 3

W14574  
ID W14574 standard; Protein; 206 AA.  
AC W14574;

DT 28-OCT-1997 (first entry)

DE Streptococcus pneumoniae PspA central region.

KW PspA; pneumococcal surface protein; vaccine; otitis media;  
KW meningitis; bacteraemia; pneumonia.

OS Streptococcus pneumoniae strain Db15.

XX Key location/Qualifiers

FT Misc-difference 50

PT /note= "unidentified amino acid"

PN WO9709994-A1.

PD 20-MAR-1997.

PF 16-SEP-1996; 96WO-US14819.

PR 15-SEP-1995; 95US-0529055.

XX (UARR-) UAB RES FOUND.

XX Briles DE, Brooks-Walter A, Crain MJ, Hollingshead S;

PI McDaniel LS, Swiatlo E, Tarl R, Yother J;

XX WPI; 1997-202002/18.

XX Streptococcus pneumoniae surface protein PspC and truncated PspA -  
PT

Streptococcus pneumoniae surface protein PspC and truncated pspA -  
used in vaccines for protecting animals against S.pneumoniae  
infection

PA (UABR-) UAB RES FOUND.  
XX  
PI Briles DE, Brooks-Wal  
PI Melhaniel's Swifts

```

XX DR WPI; 1997-202002/18.
XX PT Streptococcus pneumoniae surface protein PspC and truncated PspA -
XX PT used in vaccines for protecting animals against S.pneumoniae
XX PT infection
XX PS Example 6; Fig 13; 296pp; English.
XX PS
XX CC This sequence shows the central portion, including the C-terminus
XX CC of the alpha-helix region and some of the proline-rich region, of
XX CC pneumococcal surface protein A (PspA) of Streptococcus pneumoniae
XX CC strain Pg9163. Comparison of the N-terminal and central regions
XX CC (W14533-57 and W14562-91) of PspA polypeptides from different
XX CC pneumococcal strains can be used to divide the strains into several
XX CC families based on sequence homologies. PspA polypeptides, or
XX CC fragments of them, can be used in vaccines to protect animals
XX CC against S. pneumoniae infection and hence for the prevention of
XX CC diseases such as otitis media, meningitis, bacteraemia and pneumonia.
XX CC The sequence of the 3' half of the PspA alpha-helical region and the
XX CC immediate 5' tip of the coding sequence are likely to be the critical
XX CC sequences for predicting PspA cross-reactions and vaccine
XX CC composition.
XX SQ Sequence 190 AA;

Query Match 12.3%; Score 101.5; DB 18; Length 190;
Best Local Similarity 27.4%; Pred. No. 0.031;
Matches 32; Conservative 6; Mismatches 28; Indels 51; Gaps 5;

OY 16 APTAPSTIPGRRGSGPEIFTFDPLPEPAAAGRPASRGRKRRLVPRVVRKQI. 75
DB 111 apapapapapa-----papapapapakpapa-----pkpapapa 147

OY 76 PVGERPNAPKRLIFLLLTIVFCQILMAEGVPAPLP-PEDAPNAASLAPTPVSPVLEP 131
DB 148 papakpa-----papakpapapap-----apkpekpaekp 177

RESULT 6
W14562
ID W14562 standard; Protein; 180 AA.
XX AC W14562;
XX DT 28-OCT-1997 (first entry)
XX DE Streptococcus pneumoniae PspA central region.
XX KW PspA; pneumococcal surface protein; vaccine; otitis media;
XX KW meningitis; bacteraemia; pneumonia.
XX OS Streptococcus pneumoniae strain 0922134c.
XX PN W09709994-A1.
XX XX
XX PD 20-MAR-1997.
XX PF 16-SEP-1996; 96WO-US14819.
XX PR 15-SEP-1995; 95US-0529055.
XX PA (UABR-) UAB RES FOUND.
XX XX
XX PI Briles DE, Brooks-Walter A, Crain MJ, Hollingshead S;
XX PI McDaniel LS, Swiatlo E, Tart R, Yother J;
XX DR WPI; 1997-202002/18.
XX PT Streptococcus pneumoniae surface protein PspC and truncated PspA -
XX PT used in vaccines for protecting animals against S.pneumoniae
XX PT infection

```

```

XX PS Example 6; Fig 13; 296pp; English.
XX CC This sequence shows the central portion, including the C-terminus
XX CC of the alpha-helix region and some of the proline-rich region, of
XX CC pneumococcal surface protein A (PspA) of Streptococcus pneumoniae
XX CC strain 0922134c. Comparison of the N-terminal and central regions
XX CC (W14533-57 and W14562-91) of PspA polypeptides from different
XX CC pneumococcal strains can be used to divide the strains into several
XX CC families based on sequence homologies. PspA polypeptides, or
XX CC fragments of them, can be used in vaccines to protect animals
XX CC against S. pneumoniae infection and hence for the prevention of
XX CC diseases such as otitis media, meningitis, bacteraemia and pneumonia.
XX CC The sequence of the 3' half of the PspA alpha-helical region and the
XX CC immediate 5' tip of the coding sequence are likely to be the critical
XX CC sequences for predicting PspA cross-reactions and vaccine
XX CC composition.
XX SQ Sequence 180 AA;

```

```

Query Match 12.3%; Score 101; DB 18; Length 180;
Best Local Similarity 27.4%; Pred. No. 0.032;
Matches 31; Conservative 8; Mismatches 34; Indels 40; Gaps 4;

OY 15 QAPTAPSTIPGRRGSGPEIFTFDPLPEPAAAGRPASRGRKRRLVPRVVRKQ 74
DB 98 etpapapapapak-----papakpapapakpapak-----paekpae 139

OY 75 LPVEEPNAPKRLIFLLLTIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSP 127
DB 140 kpaep-----aek--papapekpaetpaptptp 170

```

```

RESULT 7
W14570
ID W14570 standard; Protein; 183 AA.
XX AC W14570;
XX DT 28-OCT-1997 (first entry)
XX DE Streptococcus pneumoniae PspA central region.
XX KW PspA; pneumococcal surface protein; vaccine; otitis media;
XX KW meningitis; bacteraemia; pneumonia.
XX OS Streptococcus pneumoniae strain Bg9739.
XX PN W09709994-A1.
XX XX
XX PD 20-MAR-1997.
XX PF 16-SEP-1996; 96WO-US14819.
XX PR 15-SEP-1995; 95US-0529055.
XX PA (UABR-) UAB RES FOUND.
XX XX
XX PI Briles DE, Brooks-Walter A, Crain MJ, Hollingshead S;
XX PI McDaniel LS, Swiatlo E, Tart R, Yother J;
XX DR WPI; 1997-202002/18.
XX PT Streptococcus pneumoniae surface protein PspC and truncated PspA -
XX PT used in vaccines for protecting animals against S.pneumoniae
XX PT infection
XX PS Example 6; Fig 13; 296pp; English.
XX CC This sequence shows the central portion, including the C-terminus
XX CC of the alpha-helix region and some of the proline-rich region, of
XX CC pneumococcal surface protein A (PspA) of Streptococcus pneumoniae

```

CC This protein comprises novel murine neural Mena<sup>+</sup>. Its amino acid  
CC sequence was deduced from a cDNA clone (see V02998) obtained from

Example 4: page 60-63; 77pp; English.



CC ligand dependent manner to induce gene expression can be identified  
 CC using the method of the invention. The specification describes a  
 CC method for identifying compounds that modulate a nuclear receptor.  
 CC The method comprises incubating a polypeptide containing a  
 CC ligand-binding region of nuclear receptor with test compound in  
 CC presence of second polypeptide containing the nuclear receptor-binding  
 CC part of a HMG protein, and detecting any alteration in binding between  
 CC the two polypeptides. The method is useful for identifying potential  
 CC antiproliferative, anti-inflammatory, antipsoriasis and anticancer  
 CC agents.  
 XX  
 SQ Sequence 168 AA;

Query Match 12.0%; Score 98.5; DB 21; Length 168;  
 Best Local Similarity 27.7%; Pred. No. 0.05;  
 Matches 39; Conservative 9; Mismatches 52; Indels 41; Gaps 4;  
 QY 17 PTPAPSTIPGRGSGPEITFDPLPEPAAPAGRSASGHRKRVRVLYPRVRL- 75  
 DB 33 PKKPSSEVPTPKR-----prgpkyskknwrrkrasrrsprts 74  
 QY 76 -PVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPEDAPNAASLAPTPVSPVLEPFNL 134  
 DB 75 dpcvpapahwrssflglgsf-----apipp-----ppplpqahhhrl 113  
 QY 135 TSEPSDYALDLSTFLOQHPAA 155  
 DB 114 wpppsstcaittlhtspaa 134

RESULT 12  
 Y68731  
 ID Y68731 standard; Protein; 179 AA.  
 AC Y68731;  
 DT 05-MAY-2000 (first entry)  
 DE Amino acid sequence of high mobility group (HMG)-R protein.  
 XX  
 KW Retinoic acid receptor; RAR; high mobility group protein;  
 KW HMG protein; nuclear receptor; antiproliferative; anti-inflammatory;  
 KW antipsoriasis; anticancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200001717-A2.  
 PD 13-JAN-2000.  
 XX  
 PF 28-JUN-1999; 99WO-US14801.  
 XX  
 PR 01-JUL-1998; 98US-0108298.  
 XX  
 PA (ALLR ) ALLERGAN SALES INC.  
 XX  
 PI Nagpal S, Chandratatna RA, Ghosh C;  
 XX  
 DR WPI: 2000-170999/15.  
 XX  
 PT Identifying modulators of nuclear receptors, useful e.g. as anticancer  
 PT and anti-inflammatory agents, according to their effect on binding  
 PT between the receptor and high-mobility group protein.  
 XX  
 PS Claim 24; Fig 1B; 43pp; English.  
 XX  
 CC The present sequence represents a high mobility group (HMG)-R protein.  
 CC HMG sequences that interact with retinoic acid receptors (RARs) in a  
 CC ligand dependent manner to induce gene expression can be identified  
 CC using the method of the invention. The specification describes a  
 CC method for identifying compounds that modulate a nuclear receptor.  
 CC The method comprises incubating a polypeptide containing a

CC ligand-binding region of nuclear receptor with test compound in  
 CC presence of second polypeptide containing the nuclear receptor-binding  
 CC part of a HMG protein, and detecting any alteration in binding between  
 CC the two polypeptides. The method is useful for identifying potential  
 CC antiproliferative, anti-inflammatory, antipsoriasis and anticancer  
 CC agents.  
 XX  
 SQ Sequence 179 AA;

Query Match 11.7%; Score 96.5; DB 21; Length 179;  
 Best Local Similarity 27.4%; Pred. No. 0.083;  
 Matches 40; Conservative 9; Mismatches 56; Indels 41; Gaps 4;  
 QY 12 TILQAPTAPADSTIPGRGSGPEITFDPLPEPAAPAGRSASGHRKRVRVLYPRVV 71  
 DB 39 talvsgskcsevtptkr-----prgpkyskknwrrkrasrrsp 80  
 QY 72 RRQL--PVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPEDAPNAASLAPTPVSPVL 129  
 DB 81 rrrssdpcvppaphwrssflglgsf-----apipp-----ppplpqah 119  
 QY 130 EPNLJTSEPSDYALDLSTFLOQHPAA 155  
 DB 120 hhhrlwpppsstcaittlhtspaa 145

RESULT 13  
 W25790  
 ID W25790 standard; Protein; 739 AA.  
 AC W25790;  
 DT 11-MAR-1998 (first entry)  
 DE Gene 036 product differentially expressed in colon tumour cells.  
 XX  
 KW Colon tumour; colon cancer; differential expression; gene 036;  
 KW human; diagnosis; gene therapy; tumour suppressor.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9733551-A2.  
 PD 18-SEP-1997;  
 XX  
 PF 14-MAR-1997; 97WO-US04191.  
 XX  
 PR 15-MAR-1996; 96US-0013438.  
 XX  
 PA (MILL-) MILLENNIUM PHARM.  
 XX  
 PI Shyjan AW;  
 XX  
 DR WPI: 1997-470615/43.  
 DR N-PSDB; T91708.  
 XX  
 PT Gene 036 with reduced or amplified expression in tumour cells -  
 PT used, optionally with genes 097, 030 and 056 or their protein  
 PT products, for diagnosis and treatment of colonic cancer  
 XX  
 PS Claim 38; Fig 2; 117pp; English.  
 XX  
 CC This protein is encoded by gene 036 (see T91708), which was  
 CC identified by differential display analysis as being expressed at  
 CC a higher level in normal colon tissues than in cancerous colon  
 CC tissues. Gene 036 is a candidate tumour suppressor gene. A  
 CC correlation was found between an increase in the expression level  
 CC of gene 036 and a decrease in a colon cell's tumour potential.  
 CC Hence, methods that increase the level of expression of gene 036  
 CC may inhibit or slow the progression to tumours and cancer, e.g.  
 CC colon cancer. The 036 protein is used in claimed methods for  
 CC treating a patient suffering from a disorder associated with



member - PEA3-beta. The gene was isolated from a HepG2 cell line cDNA library. The probe for the screening was prepared by reverse transcription on Hep2 mRNA followed by PCR using primers Q31770-1, to produce a probe of 170-200 bp. The screening isolated the full length sequence of the transcription factor. The gene was inserted into the expression vectors pBluescript KS and pCDV1 for expression of the protein in E.coli and COS7 cells, respectively. The ETS transcription factor has specificity for and is activated by ras. It is useful as a reagent in studies for the elucidation of the mechanism of cancer cell multiplication or polyoma virus transformation of cells.

Sequence 520 AA:  
SQ

	Query Match	11.5%	Score 94.5;	DB 16;	Length 520;
	Best Local Similarity	22.4%	Pred. No. 0.42;		
	Matches	38; Conservative	13; Mismatches	62; Indels	57; Gaps
OY	3 HRSRCHPTMTILQAPTPAPSTPIGPRRGSGPEITFDPLPAPAALPAGPSASRGHPKRS	62			
Dd	:   :   :   :   :   :   :   :   :   :   :   :   :   :				
Dd	162 hapaagvgqv--gpaaphsislep-----gpqggf-avprpbhqlpmp-----kmm	207			
OY	63 RRVLYRRVVRQLPVFERPPNPAKRLLFLLTIVFCQTLMAGEGVp-----	106			
Dd	:   :   :   :   :   :   :   :   :   :   :   :   :   :				
Dd	208 penqypseqrfqrqlsep-----chpfppqpvgvgnrpsyhrqmsepiv	252			
OY	107 --APIPEDO-----APNAASIAPTVPSPVLDPFNLTISPSDIYALD	144			
Dd	:   :   :   :   :   :   :   :   :   :   :   :				
Dd	253 paappppqkfkyghdp yhgyvgmgpgphahgfqspmgikgeprdyacd	302			

Search completed: January 30, 2001, 22:05:37  
Job time: 11886 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 18:47:25 ; Search time 1450.12 Seconds  
(without alignments)  
1246.746 Million cell updates/sec

Title: US-08-799-910-9\_COPY\_211\_468

Perfect score: 258

Sequence: 1 GTCCGGCCGACGTGCCACT.....AGCAACACCCGCCGCCCTTC 258

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estpl1:\*  
70: em\_estpl2:\*  
71: em\_estpl3:\*  
72: em\_estpl4:\*  
73: em\_estpl5:\*  
74: em\_estro1:\*  
75: em\_estro2:\*  
76: em\_estro3:\*  
77: em\_estro4:\*  
78: em\_estro5:\*  
79: em\_estro6:\*  
80: em\_estro7:\*  
81: em\_estro8:\*  
82: em\_estro9:\*  
83: em\_estro10:\*  
84: em\_estro11:\*  
85: em\_estro12:\*  
86: em\_estro13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est67:\*  
105: gb\_est68:\*  
106: gb\_est69:\*  
107: gb\_est70:\*  
108: gb\_est71:\*  
109: gb\_est72:\*  
110: gb\_est73:\*  
111: gb\_est74:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estp16.\*  
 118: em\_estp17.\*  
 119: em\_estp18.\*  
 120: em\_estp19.\*  
 121: em\_estp20.\*  
 122: em\_estp21.\*  
 123: em\_estp22.\*  
 124: em\_estp23.\*  
 125: em\_estp24.\*  
 126: em\_estp25.\*  
 127: em\_estp26.\*  
 128: em\_estp27.\*  
 129: em\_estp28.\*  
 130: em\_estp29.\*  
 131: em\_estp30.\*  
 132: em\_estp31.\*  
 133: em\_estp32.\*  
 134: em\_estp33.\*  
 135: em\_estp34.\*  
 136: em\_estp35.\*  
 137: em\_estp36.\*  
 138: em\_estp37.\*  
 139: em\_estp38.\*  
 140: em\_estp39.\*  
 141: em\_estp40.\*  
 142: em\_estp41.\*  
 143: em\_estp42.\*  
 144: em\_estp43.\*  
 145: em\_estp44.\*  
 146: em\_estp45.\*  
 147: em\_estp46.\*  
 148: em\_estp47.\*  
 149: em\_estp48.\*  
 150: em\_estp49.\*  
 151: em\_estp50.\*  
 152: em\_estp51.\*  
 153: em\_estp52.\*  
 154: em\_estp53.\*  
 155: em\_estp54.\*  
 156: em\_estp55.\*  
 157: em\_estp56.\*  
 158: em\_estp57.\*  
 159: em\_estp58.\*  
 160: em\_estp59.\*  
 161: em\_estp60.\*  
 162: em\_estp61.\*  
 163: em\_estp62.\*  
 164: em\_estp63.\*  
 165: em\_estp64.\*  
 166: em\_estp65.\*  
 167: em\_estp66.\*  
 168: em\_estp67.\*  
 169: em\_estp68.\*  
 170: em\_estp69.\*  
 171: em\_estp70.\*  
 172: em\_estp71.\*  
 173: em\_estp72.\*  
 174: em\_estp73.\*  
 175: em\_estp74.\*  
 176: em\_estp75.\*  
 177: em\_estp76.\*  
 178: em\_estp77.\*  
 179: em\_estp78.\*  
 180: em\_estp79.\*  
 181: em\_estp80.\*  
 182: em\_estp81.\*  
 183: em\_estp82.\*  
 184: em\_estp83.\*  
 185: em\_estp84.\*  
 186: em\_estp85.\*  
 187: em\_estp86.\*  
 188: em\_estp87.\*  
 189: em\_estp88.\*

190: gb\_gss25.\*  
 191: gb\_gss26.\*  
 192: gb\_gss27.\*  
 193: gb\_gss28.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	258	100.0	411	25	AI824906	AI824906 w03e06.x
2	258	100.0	544	107	BE384949	BE384949 601276496
3	258	100.0	951	107	BE383865	BE383865 601273192
4	256.4	99.4	765	106	BE275966	BE275966 601120277
5	254.4	98.6	762	110	BE621346	BE621346 601493736
6	249.4	96.7	916	137	BE882130	BE882130 601505227
7	245.4	95.1	603	106	BE273386	BE273386 601143493
8	244.6	94.8	463	7	AA410666	AA410666 zL30q09.r
9	242	93.8	585	97	AW960290	AW960290 EST372361
10	236.8	91.8	810	109	H5563592	BE563592 601334706
11	223	86.4	447	147	W52268	W52268 zc46a10.r1
12	217.2	84.2	513	11	AA743819	AA743819 OB01d04.s
13	205.4	79.6	212	142	H80073	H80073 yv80q11.r1
14	203	78.7	1070	137	BE878459	BE878459 601487866
15	197.8	76.7	504	2	AA143404	AA143404 z066e09.r
16	195.6	75.8	502	90	AW465331	AW465331 BE230018B
17	195.6	75.8	523	135	BE751344	BE751344 203178 MA
18	195.6	75.8	547	135	BE750116	BE750116 201294 MA
19	195.6	75.8	557	139	BF040186	BF040186 BP250005A
20	195.6	75.8	571	139	BF045952	BF045952 BP250003B
21	195.6	75.8	594	37	AV617905	AV617905 AV617905
22	192	74.4	346	38	AV683186	AV683186 AV683186
23	189.2	73.3	545	139	BF043065	BF043065 BP250026B
24	188.8	73.2	617	38	AV690660	AV690660 AV690660
25	181	70.2	184	143	N57203	N57203 yw91h09.r1
26	173.6	67.3	701	135	BE735733	BE735733 601304556
27	166	64.3	389	146	W39753	W39753 zc80b07.r1
28	163.8	63.5	174	137	BE939840	BE939840 RC6-UT001
29	160.8	62.3	597	37	AV594384	AV594384 AV594384
30	155	60.1	443	147	W47587	W47587 zc35b02.r1
31	155	60.1	806	139	BF025858	BF025858 601670012
32	154.4	59.8	628	96	AW916683	AW916683 EST347987
33	151.2	58.6	467	19	AI326484	AI326484 mq23c12.y
34	146.4	56.7	182	6	AA356940	AA356940 EST65572
35	145.8	56.5	150	89	AW367127	AW367127 MR0-HT015
36	144.8	56.1	487	139	BF023106	BF023106 ux04d04.y
37	143.2	55.5	471	109	BE554106	BE554106 ur38c03.y
38	139.4	54.0	1155	135	BE786915	BE786915 601477783
39	139.2	54.0	385	147	W77628	W77628 me68a09.r1
40	139.2	54.0	460	94	AW740625	AW740625 ur02h11.y
41	137	53.1	353	142	N45391	N45391 yw97d08.r1
42	136	52.7	423	2	AA122977	AA122977 mq23c12.r
43	127.6	49.5	443	147	W71619	W71619 me40d11.r1
44	126	48.8	459	26	AI876358	AI876358 uk74a05.y
45	125.4	48.6	476	12	AA833402	AA833402 ub58b04.r

## ALIGNMENTS

RESULT 1  
 AI824906/C  
 LOCUS AI824906 411 bp mRNA EST 16-DIC-1999  
 DEFINITION w03e06.x1 NC1\_CGAP\_GC6 Homo sapiens cDNA clone IMAGE:2304610 3' similar to SW:TEX1\_HUMAN P4695 RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 ; mRNA sequence.  
 ACCESSION AI824906  
 VERSION AI824906.1 GI:5445577  
 KEYWORDS EST.

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 411)  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LUNL at:  
[www-bio-llnl.gov/bbrp/image/image.html](http://bio-llnl.gov/bbrp/image/image.html)  
Insert length: 557 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 341.  
Location/Qualifiers  
1..411  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2304610"  
/clone\_lib="NCI-CCAP\_GC6"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Plasmid DNA from the normalized library  
NCI-CCAP\_GC4 was prepared, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from a pool of 5,000 clones made  
from the same library (cloneids 1257096-1258631,  
1459084-1470983, and 1475592-1476743). Subtraction by  
Bento Soares and M. Fatima Bonaldo.  
77 a 104 c 168 g 61 t 1 others  
BASE COUNT  
ORIGIN

	Query Match	100.0%;	Score 258;	DB 25;	Length 411;
	Best Local Similarity	100.0%;	Pred. No. 1e-55;		
	Matches 258;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GTCCGCGCCAGCTGCCAGTCTGAGGAACCGAACCCAGCCAAAAGCGCTTCTCTTTCTGCTG	60		
DB	282	GTCCGCGCCAGCTGCCAGTCTGAGGACCGAACCCAGCCAAAAGGCTTCTTTTCTGCTG	223		
QY	61	CTCACCATCTCTTTCTGCCAGATCTTGATGGCTGAAGAGGCTGTGCGGGCGCCCTGCCT	120		
DB	222	CTCACCATCTCTTTCTGCCAGATCTTGATGGCTGAAGAGGCTGTGCGGGCGCCCTGCCT	163		
QY	121	CCAGAGACGCGCCCTAACGCGCGATCCCTGGCGCCACCCCTGTGTGCCCGTCTCTCGAG	180		
DB	162	CCAGAGACGCGCCCTAACGCGCGATCCCTGGCGCCACCCCTGTGTGCCCGTCTCTCGAG	103		
QY	181	CCCTTTTAATCTGACTTCTCGAGCCCTCGGACTTACGCTCTTGGAGCCTACAGTCTTCTCCCTCCAG	240		
DB	102	CCCTTTTAATCTGACTTCTCGAGCCCTCGGACTTACGCTCTTGGAGCCTACAGTCTTCTCCCTCCAG	43		
QY	241	CAACACCCGGCGCGCTTTC	258		
DB	42	CAACACCCGGCGCGCTTTC	25		

RESULT	2			
BF384949				
LOCUS	HE384949	544 bp	mRNA	EST
				21-JUL-2000

DEFINITION	6012764936F1 NH_MGC_20 Homo sapiens cDNA clone IMAGE:3617543 5', mRNA sequence.
ACCESSION	BE384949
VERSION	BE384949.1 GI:9330314
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Bukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 544)
TITLE	NIH-MGC <a href="http://www.ncbi.nlm.nih.gov/MGC/">http://www.ncbi.nlm.nih.gov/MGC/</a> .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1350 Email: Robert.Strausberg@nih.gov Tissue Procurement: ATCC/DCTD/DTF CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arranged by: The I.M.A.G.E. Consortium (LMNI) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">image.llnl.gov</a> Plate: LLCM284 row: p column: 24 High quality sequence stop: 540.

```

FEATURES
  source
    high quality sequence Scop: 340.
    Location/Qualifiers
      j. 544
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:3617543"
        /clone_lib="NIH_MGC_20"
        /tissue_type="melanotic melanoma"
        /lab_host="DH10A (phage-resistant)"
        /note="organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
        EcoRI; cDNA made by oligo-dT priming. Directionally
        cloned into EcoRI/XhoI sites using the following 5'-
        adaptor: GGCACGAG(G). Size-selected >500bp for average
        insert size 1.8kb. Library constructed by Ling Hong in
        the laboratory of Gerald M. Rubin (University of
        California, Berkeley) using ZAP-cDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies)."
      98 a 225 c 131 g 90 t

```

Query Match	100.0%;	Score	258;	DB	107;	Length	544;
Best.local Similarity	100.0%;	Pred. No.	1.le-55;				
Matches	258;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Oy	1	GTCCGGCGCCAGCTGCCAGTCTGAGGAACCGAACCCAGCCAAAGGCTTCCTTTCTGCTG	60				
Db	212	GTCCGGCGCCAGCTGCCAGTCTGAGGAACCGAACCCAGCCAAAGGCTTCCTTTCTGCTG	271				
Oy	61	CTACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTCGGCGGCCCTTGCT	120				
Db	272	CTACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTCGGCGGCCCTTGCT	331				
Oy	121	CCAGAGACGCCCTTAACGGCGCATCCCTGTCGGCGCCACCCCTGTGTCCTCTCTCGAG	180				
Db	332	CCAGAGACGCCCTTAACGGCGCATCCCTGTCGGCGCCACCCCTGTGTCCTCTCTCGAG	391				
Oy	181	CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCCTGAGCACTTCCTCCAG	240				
Db	392	CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCCTGAGCACTTCCTCCAG	451				
Oy	241	CAACACCCGGCGCCTTC	258				
Db	452	CAACACCCGGCGCCTTC	469				

RESULT	3
BE383865	
LOCUS	BE383865
951 bp	mrna
EST	21-JUL-2000
DEFINITION	601273192F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3614255 5'

## mRNA sequence.

ACCESSION BE383865  
 VERSION BE383865.1 GI:9329230  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 951)  
 NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550

## REFERENCE

AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Tissue Procurement: ATCC/DCTD/DRP  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LUCM276 row: 9 column: 24  
 High quality sequence stop: 747.  
 Location/Qualifiers

## FEATURES

## source

1..951  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3614255"  
 /clone\_lib="NIH\_MGC\_20"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 210 a 324 c 253 g 164 t  
 ORIGIN

Query Match 100.0%; Score 258; DB 107; Length 951;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-55;  
 Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCCGGCGCCAGCTGCCAGTCGAGAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
 Db 210 GTCCGGCGCCAGCTGCCAGTCGAGAACCCAGCCAAAGGCTTCTTTCTGCTG 269  
 Qy 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGCTT 120  
 Db 270 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGCTT 329  
 Qy 121 CCAGAGGACGCCCTTAACGCGGCATCCCTTGGCGCCACCCCTGTGTCTCCCGCTCTCGAG 180  
 Db 330 CCAGAGGACGCCCTTAACGCGGCATCCCTTGGCGCCACCCCTGTGTCTCCCGCTCTCGAG 389  
 Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGAGTACGCTCTGGACCTTCAGACCTTCTCTCCAG 240  
 Db 390 CCCTTTAATCTGACTTCGGAGCCCTCGAGTACGCTCTGGACCTTCAGACCTTCTCTCCAG 449  
 Qy 241 CAACACCGCGCGCTTC 258  
 Db 450 CAACACCGCGCGCTTC 467

## RESULT 4

BE275966  
 LOCUS BE275966 765 bp mRNA EST 13-JUL-2000  
 DEFINITION 6011493736F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:2967534 5',  
 mRNA sequence.

## ACCESSION

BE275966  
 VERSION BE275966.1 GI:9150928  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 765)  
 NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550

## REFERENCE

AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Tissue Procurement: ATCC/DCTD/DRP  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LUCM276 row: e column: 07  
 High quality sequence start: 7  
 High quality sequence stop: 699.  
 Location/Qualifiers

## FEATURES

## source

1..765  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2967534"  
 /clone\_lib="NIH\_MGC\_20"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 151 a 284 c 202 g 128 t  
 ORIGIN

Query Match 99.4%; Score 256.4; DB 106; Length 765;  
 Best Local Similarity 99.6%; Pred. No. 2.8e-55;  
 Matches 257; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCCGGCGCCAGCTGCCAGTCGAGAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
 Db 257 GTCCGGCGCCAGCTGCCAGTCGAGAACCCAGCCAAAGGCTTCTTTCTGCTG 316  
 Qy 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGCTT 120  
 Db 317 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGCTT 376  
 Qy 121 CCAGAGGACGCCCTTAACGCGGCATCCCTTGGCGCCACCCCTGTGTCTCCCGCTCTCGAG 180  
 Db 377 CCAGAGGACGCCCTTAACGCGGCATCCCTTGGCGCCACCCCTGTGTCTCCCGCTCTCGAG 436  
 Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGAGTACGCTCTGGACCTTCAGACCTTCTCTCCAG 240  
 Db 437 CCCTTTAATCTGACTTCGGAGCCCTCGAGTACGCTCTGGACCTTCAGACCTTCTCTCCAG 496  
 Qy 241 CAACACCGCGCGCTTC 258  
 Db 497 CAACACCGCGCGCTTC 514

## RESULT 5

BE275966  
 LOCUS BE275966 762 bp mRNA EST 24-AUG-2000  
 DEFINITION 601493736F1 NIH\_MGC\_70 Homo sapiens cDNA clone IMAGE:3895736 5',  
 mRNA sequence.

BE621346  
 BE621346.1 GI:9892286  
 EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 762)  
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLC674 row: h column: 09  
 High quality sequence stop: 613.  
 Location/Qualifiers  
 1..762  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3895736"  
 /clone\_lib="NIH\_MGC\_70"  
 /tissue\_type="epithelioid carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site:1; Noli;  
 Site:2; Sali; Cloned unidirectionally. Primer: Oligo df.  
 Average insert size 1.1 kb. Library constructed by Life  
 Technologies."  
 BASE COUNT 155 a 288 c 193 g 126 t  
 ORIGIN  
 Query Match 98.6%; Score 254.4; DB J10; Length 762;  
 Best Local Similarity 99.6%; Pred. No. 8.9e-55;  
 Matches 255; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTCCGGCCCGCCAGTGCAGTCCAGAACCCGAGCCAAAGGCTTCTTCTGCTG 60  
 DB 209 CTCCGGCCCGCCAGTGCAGTCCAGAACCCGAGCCAAAGGCTTCTTCTGCTG 268  
 QY 61 CTCACCATCTGCTTCTGCGAGATCCTGATGCTGAAGAGGTGTGCCGCGCCCTGCCT 120  
 DB 269 CTCACCATCTGCTTCTGCGAGATCCTGATGCTGAAGAGGTGTGCCGCGCCCTGCCT 328  
 QY 121 CCAGAGGACGCCCTTAACGGCGCATCCCTGGCGCCACCCCTGTCCTCCCGTCTCGAG 180  
 DB 329 CCAGAGGACGCCCTTAACGGCGCATCCCTGGCGCCACCCCTGTCCTCCCGTCTCGAG 388  
 QY 181 CCTTTAATCTGACTTCGAGCCCTCGGACTTACGCTCTGGACCTCAGACATTTCTCCACG 240  
 DB 389 CCTTTAATCTGACTTCGAGCCCTCGGACTTACGCTCTGGACCTCAGACATTTCTCCACG 448  
 QY 241 CAACACCGCGCGCT 256  
 DB 449 CAACACCGCGCGCTT 464  
 RESULT 6  
 BE882130  
 LOCUS 601505227F2 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:3906823 5',  
 DEFINITION mRNA sequence.  
 BE882130  
 ACCESSION BE882130  
 VERSION BE882130.1 GI:10330906  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 603)  
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 916)  
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/  
 TITLE Unpublished (1999)  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLCM703 row: f column: 08  
 High quality sequence stop: 659.  
 Location/Qualifiers  
 1..916  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3906823"  
 /clone\_lib="NIH\_MGC\_71"  
 /tissue\_type="leiomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site:1; Noli;  
 Site:2; Sali; Cloned unidirectionally. Primer: Oligo df.  
 Average insert size 2.1 kb."  
 BASE COUNT 187 a 327 c 268 g 134 t  
 ORIGIN  
 Query Match 96.7%; Score 249.4; DB J37; Length 916;  
 Best Local Similarity 99.6%; Pred. No. 1.7e-53;  
 Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTCCGGCCCGCCAGTGCAGTCCAGAACCCGAGCCAAAGGCTTCTTCTGCTG 60  
 DB 180 GTCCGGCCCGCCAGTGCAGTCCAGAACCCGAGCCAAAGGCTTCTTCTGCTG 239  
 QY 61 CTCACCATCTGCTTCTGCGAGATCCTGATGCTGAAGAGGTGTGCCGCGCCCTGCCT 120  
 DB 240 CTCACCATCTGCTTCTGCGAGATCCTGATGCTGAAGAGGTGTGCCGCGCCCTGCCT 299  
 QY 121 CCAGAGGACGCCCTTAACGGCGCATCCCTGGCGCCACCCCTGTCCTCCCGTCTCGAG 180  
 DB 300 CCAGAGGACGCCCTTAACGGCGCATCCCTGGCGCCACCCCTGTCCTCCCGTCTCGAG 359  
 QY 181 CCTTTAATCTGACTTCGAGCCCTCGGACTTACGCTCTGGACCTCAGACATTTCTCCACG 240  
 DB 360 CCTTTAATCTGACTTCGAGCCCTCGGACTTACGCTCTGGACCTCAGACATTTCTCCACG 419  
 QY 241 CAACACCGCGC 251  
 DB 420 CAACACCGCGC 430  
 RESULT 7  
 BE273386  
 LOCUS 601143493F1 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:3507191 5',  
 DEFINITION mRNA sequence.  
 BE273386  
 ACCESSION BE273386  
 VERSION BE273386.1 GI:9148012  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 603)  
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/

TITLE  
JOURNAL  
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: DCTD/DTP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LHCMI86 row: j column: 24  
High quality sequence stop: 579.  
Location/Qualifiers

FEATURES  
source

1..603  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3507191"  
/clone\_lib="NIH\_MGC\_14"  
/tissue\_type="renal cell adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pORF7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAC(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 112 a 241 c 154 g 96 t  
ORIGIN

Query Match 95.1%; Score 245.4; DB 106; Length 603;  
Best Local Similarity 99.2%; Pred. No. 1.7e-52;  
Matches 257; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Oy 1 GTCCGGCCAGCTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCCTG 60  
Db 177 GTCCGGCCAGCTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCCTG 236  
Oy 61 CTCACCATCGTCTTCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCCGCCCT 120  
Db 237 CTCACCATCGTCTTCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCCGCCCT 296  
Oy 121 CCAGAGAGCCCTTACGGCGGATCCTGTGCGCCACCCCTGTGTCC-CCGTCCTCGA 179  
Db 297 CCAGAGAGCCCTTACGGCGGATCCTGTGCGCCACCCCTGTGTCC-CCGTCCTCGA 356  
Oy 180 CCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCGACCTCAGCACTTTCCTCCA 239  
Db 357 GCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCGACCTCAGCACTTTCCTCCA 416  
Oy 240 GCAACACCCGGCCCTTC 258  
Db 417 GCAACACCCGGCCCTTC 435

RESULT 0  
AA410666 463 bp mRNA EST 18-MAY-1997  
LOCUS zt30g09.r1 Soares ovary tumor NBH07 Homo sapiens cDNA clone  
DEFINITION IMAGE:723904 5' similar to SW:GL96\_MOUSE P46694 IMMEDIATE EARLY  
PROTEIN GLY96. ; mRNA sequence.

ACCESSION AA410666  
VERSION AA410666.1 GI:2069789  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisler, G., Jost, S.,

Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, H., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-Merck EST Project 1997  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL. contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28ml3 rev2 Et from Amersham  
High quality sequence stop: 369.  
Location/Qualifiers

FEATURES  
source

1..463  
/organism="Homo sapiens"  
/db\_xref="GDB:5935321"  
/db\_xref="taxon:9606"  
/clone="IMAGE:723904"  
/clone\_lib="Soares ovary tumor NBH07"  
/sex="Female"  
/tissue\_type="ovarian tumor"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: ovary; Vector: pT73D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTCGAGCGCGGTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."  
BASE COUNT 71 a 196 c 114 g 82 t  
ORIGIN

Query Match 94.8%; Score 244.6; DB 7; Length 463;  
Best Local Similarity 98.4%; Pred. No. 2.6e-52;  
Matches 247; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 GTCCGGCCAGCTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCCTG 60  
Db 213 GTCCGGCCAGCTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCCTG 272  
Oy 61 CTCACCATCGTCTTCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCCGCCCT 120  
Db 273 CTCACCATCGTCTTCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCCGCCCT 332  
Oy 121 CCAGAGAGCCCTTACGGCGGATCCTGTGCGCCACCCCTGTGTCC-CCGTCCTCGA 180  
Db 333 CCAGAGAGCCCTTACGGCGGATCCTGTGCGCCACCCCTGTGTCC-CCGTCCTCGA 392  
Oy 181 CCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCGACCTCAGCACTTTCCTCCA 240  
Db 393 CCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCGACCTCAGCACTTTCCTCCA 452  
Oy 241 CAACACCCGGC 251  
Db 453 CAACACCCGGC 463

RESULT 9  
AW960290

LOCUS EST372361 585 bp mRNA EST 01-JUN-2000  
DEFINITION EST372361 MAGE resequences, MAGF Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW960290  
VERSION AW960290.1 GI:8149974  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia: Eutheria; Primates: Catarrhini; Hominidae; Homo.  
 AUTHORS Hegde, P., Ol, R., Aherthy, K., Dharap, S., Gaspard, R., Gay, C., Holt  
 J.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J., and  
 Quackenbush, J.

TITLE Assessment of gene expression patterns in a model of colon tumor  
 metastasis using a 19,200 element cDNA microarray

JOURNAL Unpublished (2000)  
 COMMENT Contact: John Quackenbush  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 3528  
 Fax: 301 838 0208  
 Email: johnq@tigr.org  
 Plate: 144

Seq primer: Reverse\*

Location/Qualifiers

source 1. 585  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="MAGE resequences, MAGF"  
 /note="Vector: pBluescriptSKm"

BASE COUNT 134 a 190 c 148 g 113 t

ORIGIN

Query Match 93.8%; Score 242; DB 97; Length 585;

Best Local Similarity 96.1%; Pred. No. 1.2e-51; Indels 0; Gaps 0;  
 Matches 248; Conservative 0; Mismatches 10;

QY 1 GTCCGGCCCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTCTTTCTGCTG 60  
 DB 19 GGCCGGCCCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTCTTTCTGCTG 78  
 QY 61 CTCACCATCGTCTTCTGCCAGATCTGATGCTGAAGAGGTGTGCCGGCCCGCTGCCT 120  
 DB 79 CTCACCATCGTCTTCTGCCAGATCTGATGCTGAAGAGGTGTGCCGGCCCGCTGCCT 138  
 QY 121 CCAGAGGACGCCCTTAACCGCCGATCCCTGGCGCCACCCCTGTCTCCCGCTCTCTCAG 180  
 DB 139 CCAGAGGACGCCCTTAACCGCCGATCTGATGAGCCACCCCTGTCTCCCGCTCTCTCAG 198  
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTTCGACTACGCTCTCGACCTCAGACATTTCTCTCCAG 240  
 DB 199 CCCTTTAATCTGACTTCGTGAGCCCTCGGACTACACTCTGGACCTTCTCTCCAG 258  
 QY 241 CAACACCGCGCGGCTTC 258  
 DB 259 CAACACCGCGCGGCTTC 276

RESULT 10

BE563592  
 LOCUS BE563592 810 bp mRNA EST 15-AUG-2000  
 DEFINITION G01334706F1 NIH\_MGC\_39 Homo sapiens cDNA clone IMAGE:3688825 5',  
 mRNA sequence.  
 ACCESSION BE563592  
 VERSION BE563592.1 GI:9807312  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 810)

AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: LCM382 row: k column: 02

High quality sequence stop: 692.

FEATURES

Location/Qualifiers

source 1. 810

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="IMAGE:3688825"

/tissue\_type="adenocarcinoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI;  
 Site: 2: EcoRI; CDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAC(G). Library constructed  
 by Ling Hong in the laboratory of Gerald M. Rubin  
 (University of California, Berkeley) using ZAP-CDNA  
 synthesis kit (Stratagene) and Superscript II RT (Life  
 Technologies)."

BASE COUNT 176 a 284 c 205 g 145 t

ORIGIN

Query Match 91.8%; Score 236.8; DB 109; Length 810;

Best Local Similarity 98.8%; Pred. No. 2.6e-50; Indels 1; Gaps 1;  
 Matches 249; Conservative 0; Mismatches 2;

QY 1 GTCCGGCCCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTCTTTCTGCTG 60  
 DB 207 GTCCGGCCCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTCTTTCTGCTG 266  
 QY 61 CTCACCATCGTCTTCTGCCAGATCTGATGCTGAAGAGGTGTGCCGGCCCGCTGCCT 120  
 DB 267 CTCACCATCGTCTTCTGCCAGATCTGATGCTGAAGAGGTGTGCCGGCCCGCTGCCT 326  
 QY 121 CCAGAGGACGCCCTTAACCGCCGATCCCTGGCGCCACCCCTGTCTCCCGCTCTCTCAG 179  
 DB 327 CCAGAGGACGCCCTTAACCGCCGATCTGATGAGCCACCCCTGTCTCCCGCTCTCTCAG 386  
 QY 180 GCCCTTTAATCTGACTTCGGAGCCCTTCGACTACGCTCTGGACCTCAGACATTTCTCTCCA 239  
 DB 387 GCCCTTTAATCTGACTTCGGAGCCCTTCGACTACGCTCTGGACCTCAGACATTTCTCTCCA 446  
 QY 240 GCAACACCGGCG 251  
 DB 447 GCAACACCGGCG 458

RESULT 11

W52268  
 LOCUS W52268 447 bp mRNA EST 11-OCT-1996  
 DEFINITION zc46a10.r1 Soares\_senescent\_fibroblasts\_NBHSF Homo sapiens cDNA  
 clone IMAGE:325338 5' similar to SW:GL96\_MOUSE P46694 IMMEDIATE  
 EARLY PROTEIN GLY96. [1]; mRNA sequence.

ACCESSION W52268

VERSION W52268.1 GI:1349380

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 447)

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Paterson, J.,

Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston

R., Williamson, A., Wohlmann, P. and Wilson, R.

THE WASHU-MERCK EST PROJECT

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewatson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 3171 Std Error: 0.00

Seq primer: mob.REGA-ET

High quality sequence stop: 401.

Location/Qualifiers

#### FEATURES

source

```
1. 447
/organism="Homo sapiens"
/db_xref="taxon:9606"
/db_xref="taxon:9606"
/clone_lib="Soares_senescent_fibroblasts_NHIST"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
/notes="Vector: pf7T3D (Pharmacia) with a modified
polylinker V-TYPE: phagemid; Site_1: Not I; Site_2: Eco RI
; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pf7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo."
```

BASE COUNT 72 a 188 c 109 g 75 t 3 others

#### ORIGIN

```
Query Match 86.4%; Score 223; DB 147; Length 447;
Best Local Similarity 99.2%; Pred.No. 7.9e-47;
Matches 245; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 GTCCGGCGCCAGCTGCCAGTGGAGGAGCGGAGCGGCGCATTTTTTTTTTTTCTGCTG 60
|||||
DB 202 GTCCGGCGCCAG-TGCCAGTGGAGGAGCGGAGCGGCGCATTTTTTTTTTTTCTGCTG 260
|||||

QY 61 CTCACCATCTCTTCTGCGAGATCTGATGCTGAAGAGGGTGTCGGCGCGCCCTGCCCT 120
|||||
DB 261 CTCACCATCTCTTCTGCGAGATCTGATGCTGAAGAGGGTGTCGGCGCGCCCTGCCCT 320
|||||

QY 121 CCAGGAGGCGCCCTAAGCGCGATCCCT-GGCGCCACCCCTGTGTCCCGCTCCGCA 179
|||||
DB 321 CCAGGAGGCGCCCTAAGCGCGATCCCTGGCGCCACCCCTGTGTCCCGCTCCGCA 380
|||||

QY 180 GCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCATTTCCTCCA 239
|||||
DB 381 GCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCATTTCCTCCA 440
|||||

QY 240 GCAACAC 246
|||||
DB 441 GCAACAC 447
```

#### RESULT 12

AA743819/c

LOCUS

DEFINITION

OB01d04.s1 NCI\_CGAP\_Kid3 Homo sapiens cDNA clone IMAGE:1322407 3'

similar to SW:EX1\_HUMAN P46695 RADIATION-INDUCIBLE IMMEDIATE-EARLY

GENE IEX-1; mRNA sequence.

ACCESSION AA743819

VERSION AA743819.1 GI:2784635

KEYWORDS

EST

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 513)

REFERENCE

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Unknown library type

Insert Length: 834 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 384.

Location/Qualifiers

#### FEATURES

source

```
1. 513
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Kid3"
/lab_host="DH10B"
/notes="Organ: kidney; Vector: pf7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adapters
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pf7T3 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."
```

BASE COUNT 124 a 138 c 147 g 104 t

#### ORIGIN

```
Query Match 84.2%; Score 217.2; DB 11; Length 513;
Best Local Similarity 98.6%; Pred.No. 2.4e-45;
Matches 219; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 37 GCCAAAGGCTCTCTTCTGCTGCTCACCATCGTCTTCGCGAGATCTCGATGGCTGAA 96
|||||
DB 513 GCCAAAGGCTCTCTTCTGCTGCTCACCATCGTCTTCGCGAGATCTCGATGGCTCAA 454
|||||

QY 97 GAGGCTGTCCGGCGCCCTGCCCTCCAGAGGAGCGCCCTAACCGCATCTCCCTGGCGCCC 156
|||||
DB 453 GAGGCTGTCCGGCGCCCTGCCCTCCAGAGGAGCGCCCTAACCGCATCTCCCTGGCGCCA 394
|||||

QY 157 ACCCTGTGTCCCGCTCTCTGCTGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCT 216
|||||
DB 393 ACCCTGTGTCCCGCTCTCTGCTGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCT 334
|||||

QY 217 CTGACCTCAGCATTTCTCTCCAGACACACCCCGCGCCCTTC 258
|||||
DB 333 CTGACCTCAGCATTTCTCTCCAGACACACCCCGCGCCCTTC 292
|||||
```

#### RESULT 13

H80073

LOCUS

DEFINITION

IMAGE:249092 5' similar to SP:S33363 S33363 GLY96 PR4EIN - 1. mRNA

sequence.

ACCESSION H80073

VERSION H80073.1 GI:1058162

KEYWORDS

EST

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 212)

REFERENCE

AUTHORS

The WashU-Merk EST Project

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine



4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800

Fax: 314 286 1810

Email: estewartson.wustl.edu

High quality sequence stops: 192

Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: M13RPI

High quality sequence stop: 192.

Location/Qualifiers

## FEATURES

source

```
1. .212
/organism="Homo sapiens"
/db_xref="GDB:3866798"
/db_xref="taxon:9606"
/clone="IMAGE:249092"
/clone_lib="Soares melanocyte 2NBHM"
/sex="Male"
/tissue_type="melanocyte"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: p7T3D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGCGAGTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T3 vector
(Pharmacia). Library constructed by Dento Soares and
M.Fa'ima Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino."
34 a 81 c 49 g 44 t 4 others
```

BASE COUNT  
ORIGIN

```
Query Match 79.6%; Score 205.4; DB 142; Length 212;
Best Local Similarity 97.6%; Pred. No. 2.2e-42;
Matches 206; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 40 AAAGGCTTCTTCTCTGCTCCTACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAG 99
|||||
DB 1 AAAGAGATNNCTTTCTGCTCTACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAG 60

QY 100 COTGTCCCGCGCCCTCCCTCCAGAGACGCCCTTAACGCCGCGATCCTCGGCCCCACC 159
|||||
DB 61 GGTGTGCGCGCGCCCTCCCTCCAGAGACGCCCTTAACGCCGCGATCCTCGGCCCCACC 120

QY 160 CTTGTGTCCTCCGCTCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTG 219
|||||
DB 121 COTGTCTCCCGCGCTCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTG 180

QY 220 GACCTCAGCAGCTTCTCTCCAGACACCCCGG 250
|||||
DB 181 GACCTCAGCAGCTTCTCTCCAGACACACNCCGG 211
```

```
RESULT 14
BE878459
LOCUS BE878459 1070 bp mRNA EST 27-SEP-2000
DEFINITION 601487866F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3890248 5',
mRNA sequence.
ACCESSION BE878459
VERSION BE878459.1 GI:10327235
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1070)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
```

Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: DCTD/DTP/Gazdar  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LCM660 row: c column: 17  
High quality sequence stop: 645.

## FEATURES

source

```
1. .1070
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3890248"
/clone_lib="NIH_MGC_69"
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
276 a 324 c 302 g 168 t
```

BASE COUNT  
ORIGIN

```
Query Match 78.7%; Score 203; DB 137; Length 1070;
Best Local Similarity 98.1%; Pred. No. 1e-41;
Matches 258; Conservative 0; Mismatches 0; Indels 5; Gaps 5;

QY 1 CTCGCGCGCCAG-CTGCCAGTCGAGAACCGAACCCAGCCAGGCTTCTCTTTCTGC 58
|||||
DB 63 GTCGCGCGCCAGTCTGCCAGTCGAGAACCGAACCCAGCCAGGCTTCTCTTTCTGC 122

QY 59 TGC-TCACCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGCGCCCTG 117
|||||
DB 123 TGCATCATCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGCGCCCTG 182

QY 118 CTTCCAGAGGAC-GCCCCCTAACGCCGCGATCCTCTGCG-CCCCACCCCTGTGTCCCGCTCC 175
|||||
DB 183 CTTCCAGAGGACTGCCCTTACGCCGCGATCCTCTGCTGCCACCCCTGTGTGCCCGCTCC 242

QY 176 TCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTCGAGCTACGACTTTTCC 235
|||||
DB 243 TCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTCGAGCTACGACTTTCC 302

QY 236 TCCAGCACACCGCGCCCTTC 258
|||||
DB 303 TCCAGCACACCGCGCCCTTC 325
```

## RESULT 15

```
AA143404 504 bp mRNA EST 08-NOV-1997
LOCUS 206609.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
IMAGE:591880 5', similar to SW:GL96_MOUSE P46694 IMMEDIATE EARLY
PROTEIN GLY96.; mRNA sequence.
ACCESSION AA143404
VERSION AA143404.1 GI:1712791
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 504)
AUTHORS Hillier, L., Allon, M., Bowles, L., Dubuque, W., Geisel, C., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Ie, N., Lennon, G., Marra, M., Martin
, J., Moore, B., Schellenger, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
```

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewatson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1695 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 392.

Location/Qualifiers

FEATURES

source

1..504

/organism="Homo sapiens"

/db\_xref="GDB:4623266"

/db\_xref="taxon:9606"

/clone="IMAGE:591880"

/clone\_lib="Stratagene pancreas (#937208)"

/lab\_host="SOLR cells (kanamycin resistant)"

/note="Organ: pancreas; Vector: pBluescript SK-; Site:1:

ECORI; Site:2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. Pancreatic adenocarcinoma cell line. Average

insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor

sequence: 5' GAATTCGCCACGAC 3' -3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTTT 3' "

BASE COUNT 80 a 202 c 121 g 92 t 9 others

ORIGIN

Query Match

Best local Similarity 76.7%; Score 197.8; DB 2; Length 504;

Matches 233; Conservative 0; Mismatches 19; Indels 5; Gaps 2;

```

Oy 1  GTCCGGCGCCAGTCCGAGTCCGAGAACCC-AGCCAAAGGCTTCTCTTCTGCT 59
    |||||||  |  |||||||  |||||||  |||||||  |||||||  |||||||
Db 225 GTCCGGCGCCAGTCCGAGTCCGAGAACCC-AGCCAAAGGCTTCTCTTCTGCT 284
    |||||||  |  |||||||  |||||||  |||||||  |||||||  |||||||
Oy 60  GCTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGCCCTGCC 119
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 285 GCTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGCCCTGCC 344
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Oy 120 TCCAGAGACGCCCTAACGCCGATCCCTGGCGCCACCCCTGTGCCCCCGTCTCGA 179
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 345 TCCAGAGAGCGCCCTAACGCCGATCCCTGGCGCCACCCCTGTGCCCCCGTCTCGA 404
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Oy 180 GCCCTTTAATCTGACTTCGGAGCCCTCGG----ACTACGCTCTGGACCTCAGCACTTCC 235
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 405 GCNCTTTAATCTGACTTCGGAGCCCTTGGAACTANGCTTCTTGGAACTTAAGCACTTCC 464
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Oy 236 TCCAGCAACACCCGGCC 252
    |||||||  |||  |
Db 465 TCCAGCAACACCCGGCC 481
    |||||||  |||  |

```

Search completed: January 30, 2001, 18:47:28

Job time: 28459 sec

---



CC Surface molecule 4F2. Antisense DNA or RNA, antibodies, antibody  
 CC fragments, and peptide and non-peptide antagonists to LAT1 are useful  
 CC as anticancer agents. The present sequence encodes human LAT1, which is  
 CC specifically claimed in the present invention.

XX Sequence 4539 BP; 765 A; 1450 C; 1297 G; 998 T; 29 other;

Query Match 13.1%; Score 33.8; DB 21; Length 4539;  
 Best Local Similarity 53.0%; Pred. No. 2.6;  
 Matches 62; Conservative 2; Mismatches 53; Indels 0; Gaps 0;  
 QY 8 GCCAGCTGCCAGTCGAGGACGACGACCCAGCCGAAAGGCTTCCTTCCTGCTGCTCAACCA 67  
 DB 556 gcccggtgcccgagggagcagcgaagctcgtggcctggytntggtgctgctcaegg 615  
 QY 68 TCCCTCTCCACACCTCATGCTCAAGACGCTCTCCCGCGCCCTCCCTCCACG 124  
 DB 616 ccylgaactgclacagcgtgaaggcngcnacncggatccaggatgcltgcgcgcg 672

## RESULT 13

Q70728/c  
 ID Q70728 standard; cDNA; 1120 BP.

XX Q70728;

XX 23-MAR-1995 (first entry)

DE TATA-binding protein-associated factor dTAFII40 cDNA.

XX TATA-binding protein associated factor; dTAFII40; ss; screening;  
 KW diagnostic; therapeutic; gene transcription regulation.

XX Drosophila.

XX Key Location/Qualifiers  
 FH 80..913  
 FT /\*tag= a  
 FT /\*tag= a

PN W09417087-A.

XX 04-AUG-1994.

XX 28-JAN-1994; 94WO-US01114.

XX 28-JAN-1993; 93US-0013412.

PR 30-JUN-1993; 93US-0087119.

XX (REGC ) UNIV CALIFORNIA.

XX Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;  
 PI Tjian R, Wang E, Weinzierl ROJ;

XX WPI: 1994-264019/32.  
 DR P-PSDB; R56490.

XX TATA-binding protein associated protein factors - and  
 PT corresponding nucleotide sequence and deriv. antibodies, useful  
 PT in screening, diagnostics and therapeutics

XX Disclosure; Page 89-92; 180pp; English.

XX The TATA-binding protein associated factor dTAFII40 (including  
 CC specific antibodies and fusion products) are used in drug screening,  
 CC diagnostics and therapeutics. They are used in the development of  
 CC specific biochemical assays for screening compounds that agonise or  
 CC antagonise selected transcription factors involved in regulating  
 CC gene expression associated with human pathology.

XX Sequence 1120 BP; 307 A; 292 C; 311 G; 210 T; 0 other;

Query Match 13.0%; Score 33.6; DB 15; Length 1120;  
 Best Local Similarity 51.0%; Pred. No. 2.2;  
 Matches 104; Conservative 0; Mismatches 99; Indels 1; Gaps 1;

QY 34 CCAGCCAAAGGCTTCTCTTCTGCTGCATCACCATCGTCTTCTGCCAGATCTGATGGCT 93  
 DB 866 CCACTCCAAAGCCAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 808

QY 94 GAAGAGGTGTCCGGCGCCCTGCTCCAGAGGAGCCCTTAACCCGCATCCCTTGGCG 153  
 DB 807 GAACCGGATGCCCGCAGCATGCTGCCACGCGCGCCGATCGTGTCCACCTCCATCTTG 748

QY 154 CCCACCCCTGTCTCCCGCTCTCGATCTCTTTAATCTGACTTGGAGCCCTGGACTAC 213  
 DB 747 AGATCTCCCGCGCGCGCTGCTCTCGCTCTTAACCTCTCTGACCCAGCCGCCCCAGAT 688

QY 214 GCTCTGGACCTCAGCATTTCCTC 237

DB 687 CGCGCGGAGCTGCCACCGTTTTC 664

## RESULT 14

T42214/c  
 ID T42214 standard; cDNA; 1120 BP.

XX T42214;

XX 27-JAN-1997 (first entry)

DE Drosophila TATA-binding protein associated factor dTAFII60 gene.

XX Drosophila; TATA-binding protein; TBP associated factor; TFIIID;  
 KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;  
 KW holoenzyme; lambda-gtl; expression library; ds.

OS Drosophila melanogaster.

XX Key Location/Qualifiers  
 FH 80..916  
 FT /\*tag= a  
 FT /\*tag= a  
 FT /\*product= Drosophila TAFII60

XX US5534410-A.

XX 09-JUL-1996.

XX 28-JAN-1993; 93US-0013412.

XX 28-JAN-1994; 94US-0188582.

PR 28-JAN-1993; 93US-0013412.

PR 30-JUN-1993; 93US-0087119.

XX (REGC ) UNIV CALIFORNIA.

XX Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;  
 PI Tjian R, Wang E, Weinzierl ROJ;

XX WPI: 1996-333245/33.  
 DR P-PSDB; T42214.

XX Screen for cpds. that bind human TATA-binding protein associated  
 PT factor - by testing ability to bind to polypeptide fragments of the  
 PT factor, useful as (ant)agonists of transcription factors involved in  
 PT disease.

XX Examples; Column 61-64; 86pp; English.

XX This is the nucleotide sequence encoding the Drosophila TATA-binding  
 CC protein (TBP) associated factor (TAF) designated TAFII60. The protein  
 CC is a component of the TFIIID fraction required for reconstituting RNA  
 CC polymerase II in vitro transcription activity. The encoded protein  
 CC has an estimated mol. wt. of 60 kD by SDS-PAGE.  
 CC The invention relates to purified proteins involved in transcription



	Query Match	14.08;	Score 36.2;	DB 21;	Length 1000;
	Best Local Similarity	44.72;	Pred. No. 0.42;		
	Matches 68;	Conservative	0;	Mismatches	84;
				Indels	0;
	Gaps	0;			
OY	106	CCGCGGCCCTGCCTCCAGAGAGCGCCCTAACGGCCCATCCTTGGCGCCACCCCCTGTG	165		
IB	803	CCTTCCCCCGCCCCCGCCNNGGNCCTCTCCCGCCNCCTTCGCCGNNCCGCCGCCNCCN	744		

V69706 standard; DNA; 170 BP.  
V69706:  
08-FEB-1999 (first entry)  
Nucleotide sequence of intron 3 of iENTP gene.  
Equilibrative nucleoside transport protein; iENTP; NBMPK; transport;  
nitrobenzylmercapto-purine riboside; antiviral; antitumour; screening;  
inhibitor; drug; knockout mice; cancer; chemotherapy; gene therapy;  
adenosine deaminase; factor VIII; intron; ds.  
Homo sapiens.  
WO9846749-A1.  
22-OCT-1998.  
10-APR-1998; 98WO-US07283.  
09-APR-1998; 98US-0058389.  
11-APR-1997; 97US-0838645.  
(SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
Belt JA, Crawford CR, Patel DH;  
WPI: 1998-594576/50.  
New equilibrative nucleoside transport protein insensitive to  
nitrobenzylthio-purine riboside - useful for, e.g. identifying  
specific therapeutic nucleoside analogues and in gene therapy to  
protect transduced cells against ablative chemotherapy  
Disclosure: Page 88; 114pp; English.  
Sequences V69704 to V69715 represent introns of the gene encoding an  
equilibrative nucleoside transport protein (iENTP), which is insensitive  
to nitrobenzylmercapto-purine riboside (NBMPR). Cells transfected with a  
construct containing the iENTP nucleic acid can be used to produce the  
protein recombinantly. iENTP is used to identify specific ligands  
(particularly antiviral and antitumour nucleoside analogues that are  
preferentially transported into cells) and to raise antibodies. Cells in  
which iENTP provides all available transport activity are used: (a) to  
identify permeants of iENTP and (b) to screen specific inhibitors of  
iENTP (potential drugs). Fragments of the iENTP nucleic acid are used,  
as probes, primers, antisense molecules, and ribozymes for therapy or  
diagnosis, and knockout mice in which both alleles encoding iENTP  
contain an inactivating defect are also useful for drug screening. Cells  
that have been transduced with iENTP nucleic acid ex vivo are used  
particularly, for cancer chemotherapy. Vectors in which the iENTP gene  
is linked to a heterologous gene (e.g. encoding adenosine deaminase and  
factor VIII) can be used for gene therapy.  
Sequence 170 BP; 18 A; 77 C; 42 G; 33 T; 0 other;

242235 standard; cDNA; 1526 BP.  
242235:  
31-JAN-2000 (first entry)  
Human normal bladder tissue cDNA derived EST 114.  
Human; bladder; treatment; EST; expressed sequence tag; cytosolic;  
cancer; gene therapy; ss.  
Homo sapiens.  
DE19818620-A1.  
28-OCT-1999.  
21-APR-1998; 98DE-1018620.  
21-APR-1998; 98DE-1018620.  
(META-) METAGEN GES GENOMFORSCHUNG MBH.  
Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
WPI: 1999-602416/52.  
P-PSDB; Y60556, Y60557, Y60558.  
New polypeptides and their nucleic acids, useful for treatment of  
bladder tumour and identification of therapeutic agents  
Claim 3; Page 234; 366pp; German.  
This invention describes novel polypeptide fragment sequences (I) and  
their encoding nucleic acids (II) which are highly expressed in normal  
bladder tissue and have cytostatic activity. (II) are used for  
recombinant expression of (I) and to isolate complete genes. (I) are  
used to identify agents suitable for the treatment of bladder tumours, to  
directly treat this form of cancer (including expression from gene  
therapy vectors), or are used in a preparation for cancer treatment. (I)  
is also used for the generation of specific antibodies. (II) are  
identified by assembling ESTs (expressed sequence tags) from a particular  
tissue type before comparison of expression patterns. This allows a  
significantly longer fragment of the gene to be revealed, and therefore  
reduces the number of failures because of ESTs from different libraries  
representing different parts of the same unknown gene distorting the  
estimated frequency of occurrence in a particular tissue. 242122-242248  
represent EST fragments derived from a human normal bladder tissue cDNA  
library which encode the protein fragments represented in Y60329-Y60591.  
Sequence 1526 BP; 365 A; 523 C; 437 G; 201 T; 0 other;

Query Match 14.1%; Score 36.4; DB 20; Length 1526;  
Best Local Similarity 50.4%; Pred. No. 0.41;  
Matches 114; Conservative 0; Mismatches 111; Indels 1; Gaps 1;  
Qy 2 TCCGGCCGCGCTGCCAGTCCGAGGACCGAACCCAGCAAAAGGCTTCTTTCTGCTGC 61  
Db 240 TGCTGCTGCAATTGCTGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181  
Qy 62 TCACCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTCGCGGCCCTTCTGCTC 121  
Db 180 TGTGCTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122  
Qy 122 CAGAGGACGCCCTTAACGCCGCTATCCCTGGCGGCCACCCCTGTGTCCTCCCTCTCTG 181  
Db 121 CCCTCAACGCTCAAGTCCGCCCCACGGATGGCGGCCGCCGCCGCCGCCGCCGCC 62  
Qy 182 CTTTATCTGACTTCGGAGCCCTTCGGAATACGCTCTGGACCTCAG 227  
Db 61 GGCGGCTCGGGCCCGGGGCTCCCGGCTCCCGGCTCCCGGCTGCTACTCCG 16





```
DT 15-FEB-1999 (first entry)
XX EST clone BY66.
XX
XX Human: secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibitor; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW gene therapy; ss.
XX
XX Homo sapiens.
OS
XX W09845436-A2.
XX
XX 15-OCT-1998.
PD
XX 10-APR-1998: 98WO-US06955.
XX
XX 10-APR-1997: 97US-0838821.
XX
XX (CEMY ) GENETICS INST INC.
PA
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
PI
XX WPI: 1999-070077/06.
XX
XX New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
XX
XX Claim 1: Page 164; 618pp; English.
XX
XX The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
XX
XX Sequence 193 BP; 35 A; 39 C; 68 G; 51 T; 0 other;
SQ

Query Match 32.6%; Score 84; DB 20; Length 193;
Best Local Similarity 100.0%; Pred. No. 4.2e-14;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 CTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTCGACCTCAGCATTTC 234
DB 193 CTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTCGACCTCAGCATTTC 134
QY 235 CTCGAGCAACACCCGCGCCTTC 258
DB 133 CTCGAGCAACACCCGCGCCTTC 110

RESULT 4
Q77534
ID Q77534 standard; DNA; 297 BP.
XX
XX Q77534:
XX
XX 23-SEP-1994 (first entry)
XX
XX Human genome fragment.
DE
XX Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.
```

```
XX Homo sapiens.
OS
XX W09401548-A.
XX
XX 20-JAN-1994.
PD
XX 13-JUL-1993: 93WO-GB01467.
XX
XX 13-JUL-1992: 92GB-0014857.
XX
XX (MEDI-) MEDICAL RES COUNCIL.
PA
XX Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
PI Sibson DR, Starkey M;
XX
XX WPI: 1994-035056/04.
XX
XX New nucleic acid fragment encoding gene products - can be used
PT for genetic analysis and mapping
PT
XX Claim 1: Page 575-576; 616pp; English.
XX
XX Human nucleic acid fragments, isolated from brain, adrenal tissue,
CC the placenta or bone marrow comprise any of: (A) a sequence
CC selected from (Q76401-Q77613), (B) an allelic variation of a
CC sequence as described in (A), or (C) a sequence complementary
CC to (A) or (B).
CC Preferred sequences exhibit no more than 90% homology to a human
CC sequence known per se.
XX
XX Sequence 297 BP; 66 A; 117 C; 61 G; 52 T; 1 other;
SQ

Query Match 17.1%; Score 44; DB 15; Length 297;
Best Local Similarity 77.9%; Pred. No. 0.0026;
Matches 53; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2 TCCGGCGCAGCTGCCAGTCGAGAACCGAACCCAGCCAAAAGGCTTCTCTTCTGCTGC 61
DB 225 tccggcgccagttgcagtaaggagacaaacctgccagaagctcttccgggttga 284
QY 62 TCACCATC 69
DB 285 tcgccatc 292

RESULT 5
V69699
ID V69699 standard; DNA; 6354 BP.
XX
XX V69699:
XX
XX 08-FEB-1999 (first entry)
XX
XX 5' UTR sequence of iENTP DNA.
XX
XX Equilibrative nucleoside transport protein; iENTP; NBMPR; transport;
KW nitrobenzylmercaptopyrine riboside; antiviral; antitumour; screening;
KW inhibitor; drug; knockout mice; cancer; chemotherapy; gene therapy;
KW adenosine deaminase; factor VIII; UTR; untranslated region; ds.
XX
XX Homo sapiens.
OS
XX W09846749-A1.
XX
XX 22-OCT-1998.
PD
XX
XX 10-APR-1998: 98WO-US07283.
XX
XX 09-APR-1998: 98US-0058389.
XX
XX 11-APR-1997: 97US-0838845.
XX
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 21:29:29 ; Search time 77.57 seconds

(without alignments)  
1249.465 Million cell updates/sec

Title: US-08-799-910-9\_COPY\_211\_468

Perfect score: 258

Sequence: 1 GTCCGGCCGACGTCAGT.....AGCAACACCGCGCCGCTTC 258

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: -480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_36:\*

- 1: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1980.DAT:\*
- 2: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1981.DAT:\*
- 3: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1982.DAT:\*
- 4: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1983.DAT:\*
- 5: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1984.DAT:\*
- 6: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1985.DAT:\*
- 7: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1986.DAT:\*
- 8: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1987.DAT:\*
- 9: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1988.DAT:\*
- 10: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1989.DAT:\*
- 11: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1990.DAT:\*
- 12: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1991.DAT:\*
- 13: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1992.DAT:\*
- 14: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1993.DAT:\*
- 15: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1994.DAT:\*
- 16: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1995.DAT:\*
- 17: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1996.DAT:\*
- 18: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1997.DAT:\*
- 19: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1998.DAT:\*
- 20: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1999.DAT:\*
- 21: /cgn2\_2/gcgdata/geneseq/geneseqn/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	258	100.0	1228	18	T94471		Human Fchd605 gene
2	258	100.0	1228	21	T50711		Nucleotide sequenc
3	84	32.6	193	20	V89271		EST clone Bf66. H
4	44	17.1	297	15	Q77534		Human genome fragm
5	38.6	15.0	6354	19	V69699		5' UTR sequence of
6	38.2	14.8	4228	21	Z51683		Human cyclic nucle
7	37.6	14.6	170	19	V69706		Nucleotide sequenc
8	36.4	14.1	1526	20	Z42235		Human normal bladd
9	36.2	14.0	1000	21	A02484		Human colon cancer
10	34.8	13.5	2953	19	V62462		Rat transcription
11	34.2	13.3	3690	13	Q30849		Type III procollag
12	33.8	13.1	4539	21	A08372		Human L-type amino

C 13	33.6	13.0	1120	15	O70728	TATA-binding prote
C 14	33.6	13.0	1120	17	T42214	Drosophila TATA-bi
C 15	33.6	13.0	1120	18	T79601	TATA-binding prote
C 16	33.6	13.0	3765	13	Q31890	FMR-1 gene for det
C 17	33.6	13.0	5460	17	T16508	Vector PAC3A1 cont
C 18	33	12.8	1072	13	Q24285	Epo:IL-3 Flex, rec
C 19	33	12.8	11722	19	V34455	Human MHC class I
C 20	33	12.8	114955	20	X53491	Human adenosine A1
C 21	32.8	12.7	513	20	X91237	T. gondii immunoge
C 22	32.8	12.7	514	20	X91400	T. gondii immunoge
C 23	32.8	12.7	1450	21	Z47131	Pseudomonas aerugi
C 24	32.8	12.7	8169	19	V26609	Actinomadura hibis
C 25	32.4	12.6	201	8	N70194	Signal portion of
C 26	32.4	12.6	201	8	N70195	Streptomyces prote
C 27	32.4	12.6	2681	20	X58429	Thermophilus therm
C 28	32.2	12.5	1524	21	A08393	Human L-type amino
C 29	32	12.4	742	21	Z45675	cDNA sequence of a
C 30	32	12.4	1459	21	A02528	Human colon cancer
C 31	32	12.4	3410	19	V61201	Full length cDNA s
C 32	32	12.4	3410	19	V58586	Prostate tumour sp
C 33	32	12.4	3410	21	A06349	Human immunogenic
C 34	32	12.4	34094	20	Z30463	Complete nucleotid
C 35	31.8	12.3	599	19	V43880	Mycobacterial meth
C 36	31.8	12.3	1321	21	A37128	Human PRO1781 (UNQ
C 37	31.6	12.2	1586	19	V34287	Human secreted pro
C 38	31.6	12.2	1907	19	V34314	Human secreted pro
C 39	31.6	12.2	2074	19	V32932	Human cyclin-depen
C 40	31.6	12.2	2074	19	V32933	Human cyclin-depen
C 41	31.4	12.2	595	14	Q50041	ICP34.5 fragment.
C 42	31.4	12.2	1335	13	Q23295	HSV-1 (F) ICP34.5
C 43	31.4	12.2	13842	21	Z87297	S. venezuelae mæc
C 44	31.4	12.2	36778	21	Z87318	S. venezuelae pik
C 45	31.4	12.2	37948	21	Z87285	S. venezuelae pik

#### ALIGNMENTS

RESULT 1

T94471

ID T94471 standard; cDNA; 1228 BP.

AC T94471;

DT 03-MAR-1998 (first entry)

XX Human Fchd605 gene differentially regulated in monocytes.

DE Fchd602 gene; differential expression; monocyte; human;

KW foam cell; cardiovascular disease; atherosclerosis; ischaemia;

KW reperfusion; hypertension; restenosis; arterial inflammation;

KW therapy; diagnosis; drug screening; marker; ss.

XX Homo sapiens.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 1..471

FT /\*tag= a

PN WO9730065-A1.

XX 21-AUG-1997.

PD 21-AUG-1997.

XX 14-FEB-1997; 97WO-US02291.

PF 14-FEB-1997; 97US-0799910.

PR 13-FEB-1997; 96US-0011787.

PR 16-FEB-1996; 96US-0011787.

XX (MILL-) MILLENNIUM PHARM INC.

PA Falb DA;

PI Falb DA;

XX WPI; 1997-424966/39.

DR

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 10:53:09 ; Search time 1450.12 Seconds  
(without alignments)  
5934.126 Million cell updates/sec

Title: US-08-799-910-9  
Perfect score: 1228  
Sequence: 1 ATCTGTCACCTCCGAGCTG.....AAAAAATAAACTCGAG 1228

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues  
Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estpl1:\*  
70: em\_estpl2:\*  
71: em\_estpl3:\*  
72: em\_estpl4:\*  
73: em\_estpl5:\*  
74: em\_estrol:\*  
75: em\_estrol2:\*  
76: em\_estrol3:\*  
77: em\_estrol4:\*  
78: em\_estrol5:\*  
79: em\_estrol6:\*  
80: em\_estrol7:\*  
81: em\_estrol8:\*  
82: em\_estrol9:\*  
83: em\_estrol10:\*  
84: em\_estrol11:\*  
85: em\_estrol12:\*  
86: em\_estrol13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est58:\*  
105: gb\_est59:\*  
106: gb\_est60:\*  
107: gb\_est61:\*  
108: gb\_est62:\*  
109: gb\_est63:\*  
110: gb\_est64:\*  
111: gb\_est65:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estpl6:\*  
118: em\_estpl7:\*  
119: em\_estpl8:\*  
120: em\_estpl9:\*  
121: em\_estpl10:\*  
122: em\_estpl11:\*  
123: em\_estpl12:\*  
124: em\_estpl13:\*  
125: em\_estpl14:\*  
126: em\_estpl15:\*  
127: em\_estpl16:\*  
128: em\_estpl17:\*  
129: em\_estpl18:\*  
130: em\_estpl19:\*  
131: em\_estpl20:\*  
132: em\_estpl21:\*  
133: em\_estpl22:\*  
134: em\_estpl23:\*  
135: em\_estpl24:\*  
136: em\_estpl25:\*  
137: em\_estpl26:\*  
138: em\_estpl27:\*  
139: em\_estpl28:\*  
140: em\_estpl29:\*  
141: em\_estpl30:\*  
142: em\_estpl31:\*  
143: em\_estpl32:\*  
144: em\_estpl33:\*  
145: em\_estpl34:\*  
146: em\_estpl35:\*  
147: em\_estpl36:\*  
148: em\_estpl37:\*  
149: em\_estpl38:\*  
150: em\_estpl39:\*  
151: em\_estpl40:\*  
152: em\_estpl41:\*  
153: em\_estpl42:\*  
154: em\_estpl43:\*  
155: em\_estpl44:\*  
156: em\_estpl45:\*  
157: em\_estpl46:\*  
158: em\_estpl47:\*  
159: em\_estpl48:\*  
160: em\_estpl49:\*  
161: em\_estpl50:\*  
162: em\_estpl51:\*  
163: em\_estpl52:\*  
164: em\_estpl53:\*  
165: em\_estpl54:\*  
166: em\_estpl55:\*  
167: em\_estpl56:\*  
168: em\_estpl57:\*  
169: em\_estpl58:\*  
170: em\_estpl59:\*  
171: em\_estpl60:\*  
172: em\_estpl61:\*  
173: em\_estpl62:\*  
174: em\_estpl63:\*  
175: em\_estpl64:\*  
176: em\_estpl65:\*  
177: em\_estpl66:\*  
178: em\_estpl67:\*  
179: em\_estpl68:\*  
180: em\_estpl69:\*  
181: em\_estpl70:\*  
182: em\_estpl71:\*  
183: em\_estpl72:\*  
184: em\_estpl73:\*  
185: em\_estpl74:\*  
186: em\_estpl75:\*  
187: em\_estpl76:\*  
188: em\_estpl77:\*  
189: em\_estpl78:\*

190: gb\_gss25:\*  
191: gb\_gss26:\*  
192: gb\_gss27:\*  
193: gb\_gss28:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	748.2	60.9	951	107	BE383865	BE383865 601273192
2	621.4	50.6	765	106	BE275966	BE275966 601120727
3	604.2	49.2	762	110	BE621346	BE621346 601493736
4	590	48.0	916	137	BE882130	BE882130 601505227
5	589.4	48.0	603	106	BE273386	BE273386 601143493
6	588	47.9	810	109	BE563592	BE563592 601334706
7	585	47.6	585	137	BE910077	BE910077 601502143
8	581.2	47.3	586	89	AW340002	AW340002 hc91d03.x
9	579.6	47.2	598	93	AW662348	AW662348 hi25c07.x
10	567.2	46.2	614	110	BE621600	BE621600 601493736
11	561.6	45.7	588	97	AW956285	AW956285 EST368355
12	557	45.4	617	38	AV690660	AV690660 AV690660
13	544.8	44.4	548	88	AW263252	AW263252 xn79q02.x
14	540.8	44.0	1070	137	BE878459	BE878459 601487866
15	519.4	42.3	521	25	A1800828	A1800828 wg13f12.x
16	517.4	42.1	544	107	BE384949	BE384949 601276496
17	515.6	42.0	523	18	A1302813	A1302813 qn58g09.x
18	514.8	41.9	585	97	AW960290	AW960290 EST372361
19	513.4	41.8	518	12	AA806234	AA806234 oe29h02.s
20	512	41.7	513	40	AW150706	AW150706 xg38a10.x
21	506.2	41.2	524	22	A1566501	A1566501 tr60f07.x
22	505.4	41.2	509	9	AA582942	AA582942 nn80c01.s
23	503.4	41.0	521	23	A1685453	A1685453 tr88d12.x
24	501.6	40.8	510	89	AW337920	AW337920 xw63a07.x
25	496.4	40.4	553	142	N32077	N32077 yw97d08.s1
26	494.8	40.3	1155	135	BE786915	BE786915 601477783
27	487.6	39.7	548	96	AW887044	AW887044 RC1-OT008
28	486.8	39.6	535	25	A1802925	A1802925 tj35a07.x
29	486.6	39.6	494	15	A1022951	A1022951 ow66g09.s
30	482.4	39.3	484	19	A1336188	A1336188 ql44c12.x
31	481.4	39.2	491	39	AW081284	AW081284 xc30c01.x
32	481	39.2	492	10	AA643850	AA643850 np26f08.s
33	478.4	39.0	480	96	AW874420	AW874420 hq04a02.x
34	472.4	38.5	496	16	A1160053	A1160053 qc08a02.x
35	472	38.4	475	8	AA512938	AA512938 nh91b09.s
36	470.8	38.3	487	23	A1687081	A1687081 tp92h08.x
37	467	38.0	701	135	BE735733	BE735733 601304556
38	464.6	37.8	483	16	A1091958	A1091958 qn59d10.s
39	461.4	37.6	466	39	AW051587	AW051587 yw87h05.x
40	458.4	37.3	460	17	A1185199	A1185199 qe35h09.s
41	458.4	37.3	460	24	A1738521	A1738521 w133a06.x
42	458.4	37.3	463	13	AA884985	AA884985 am35e07.s
43	453	36.9	461	12	AA812286	AA812286 nr82f07.s
44	451.6	36.8	459	147	W60982	W60982 2c98h09.s1
45	450	36.6	469	24	A1761431	A1761431 wq65e08.x

ALIGNMENTS

RESULT 1  
BE383865  
LOCUS 601273192F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:3614255 5',  
DEFINITION BE383865 951 bp mRNA ESW 21-JUL-2000  
ACCESSTION BE383865  
VERSION BE383865.1 GI:9329230  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 951)  
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: ATCC/DC/DTP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LICM276 row: 9 column: 24  
High quality sequence stop: 747.  
Location/Qualifiers  
1..951  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3614235"  
/clone\_lib="NIH-MGC\_20"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dr priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
210 a 324 c 253 g 164 t

BASE COUNT 210 a 324 c 253 g 164 t  
ORIGIN

Query Match 60.9%; Score 748.2; DB 107; Length 951;  
Best Local Similarity 91.5%; Pred. No. 4.8e-180;  
Matches 848; Conservative 0; Mismatches 73; Indels 6; Gaps 5;

Qy 2 TGTGTCACTTCGCCAGTCCACCGACCATCACCATCTCGCAGGCCCGCCGCGCCG 61  
Db 1 TGTGTCACTTCGCCAGTCCACCGACCATCACCATCTCGCAGGCCCGCCGCGCCG 60  
Qy 62 CTTCCACCATCCCGGACCCCGGGGCTCGGTCTGAGATCTTCACCTTCGACCTC 121  
Db 61 CTTCCACCATCCCGGACCCCGGGGCTCGGTCTGAGATCTTCACCTTCGACCTC 120  
Qy 122 TCCCGGAGCCGCGAGCGGCCCTCGCGGCGCCCGCCAGCCCTCTCGCGGCGACCGAAAGC 181  
Db 121 TCCCGGAGCCGCGAGCGGCCCTCGCGGCGCCCGCCAGCCCTCTCGCGGCGACCGAAAGC 180  
Qy 182 GCAGCCCGCCAGGCTCTCTACCTTCGAGTGTCTCGCGGCCCGCAGTGCAGTCGAGGAACCGA 241  
Db 181 GCAGCCCGCCAGGCTCTCTACCTTCGAGTGTCTCGCGGCCCGCAGTGCAGTCGAGGAACCGA 240  
Qy 242 ACCCAGCCAAAGGCTTCTCTTCTGCTCTCCTCACCATCTCTTCTGCCAGATCTTGATGG 301  
Db 241 ACCCAGCCAAAGGCTTCTCTTCTGCTCTCCTCACCATCTCTTCTGCCAGATCTTGATGG 300  
Qy 302 CTGAAGAGGCTGTGCGGGGCTCTCGCTCCAGAGGAGCGCCCTTACGCGCGCATCCCTGG 361  
Db 301 CTGAAGAGGCTGTGCGGGGCTCTCGCTCCAGAGGAGCGCCCTTACGCGCGCATCCCTGG 360  
Qy 362 CGCCCGCCCGCTGTCTCCCGCTCTCGAGCCCTTTTAACTGACTTCGAGCCCTTCGAGCT 421  
Db 361 CGCCCGCCCGCTGTCTCCCGCTCTCGAGCCCTTTTAACTGACTTCGAGCCCTTCGAGCT 420  
Qy 422 ACCTCTTGACCTCAGACTTCTCTCCTCAGACACCGCGCGCTTCTTACTGTGACTCC 481  
Db 421 ACCTCTTGACCTCAGACTTCTCTCCTCAGACACCGCGCGCTTCTTACTGTGACTCC 480

Qy 482 CCGCACTCCCAAAAGAAATCCGAAACACCAAGAAACACCAAGGCTTACTTGTGCGC 541  
Db 481 CCGCACTCCCAAAAGAAATCCGAAACACCAAGAAACACCAAGGCTTACTTGTGCGC 540  
Qy 542 GAGAGCGTATCCCAACTTGGGACTTCCGAGGCACTTGAATCAACACACTACAGCCGAG 601  
Db 541 GAGAGCGTATCCCAACTTGGGACTTCCGAGGCACTTGAATCAACACACTACAGCCGAG 599  
Qy 602 ACGCCACCGCGTCTTGAGCGGGGCGGAGCGGCACAGAGACCGGCGCATAGAGCCG 661  
Db 600 ACGCCACCGCGTCTTGAGCGGGGCGGAGCGGCACAGAGACCGGCGCATAGAGCCG 659  
Qy 662 AGGCACAGCCGAGC-TGGGGCTAGGCCCGTGGGAAGGAGAGCGCTTAAATTTATCT 720  
Db 660 AGGCACAGCCGAGC-TGGGGCTAGGCCCGTGGGAAGGAGAGCGCTTAAATTTATCT 719  
Qy 721 TATTGCTCTTAATTAATTAATTAATG--TATTTATGACGTCCTCCTAGGTAGTGAGAT 778  
Db 720 TATAGCTCTTAATTAATTAATTAATG--TATTTATGACGTCCTCCTAGGTAGTGAGAT 779  
Qy 779 GTGTACGTAATTAATTAATTAATTAATGATGTTTCCCTCTGCTGCTGAAT 838  
Db 780 GTGTACGTAATTAATTAATTAATTAATGATGTTTCCCTCTGCTGCTGAAT 837  
Qy 839 GCAGTCTCTTTGGTATTTTATGAGCTTTGTGGGACTGCTGGAAGCAGACACCTGGAAT 898  
Db 838 AGGCACAGGTAATTAATTAATTAATGAGCTTTGTGGGACTGCTGGAAGCAGACACCTGGAAT 897  
Qy 899 CGGCAAAAGTAGGAGAGAAATGGGGA 925  
Db 898 AGGGAAGACCGCGGACCATGAGATGGGA 924

RESULT 2  
BE275966 765 bp mRNA 13-JUL-2000  
LOCUS BE275966 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:2967534 5',  
DEFINITION mRNA sequence.  
ACCESSION BE275966  
VERSION BE275966.1 GI:9150928  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 765)  
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: ATCC/DC/DTP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LICM272 row: e column: 07  
High quality sequence start: 7  
High quality sequence stop: 699.  
Location/Qualifiers  
1..765  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2967534"  
/clone\_lib="NIH\_MGC\_20"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dr priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'

FEATURES  
source

adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by King Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

HASH COUNT 151 a 284 c 202 g 128 t

ORIGIN

Query Match 50.6%; Score 621.4; DB 106; Length 765;  
 Best Local Similarity 99.4%; Pred. No. 9.2e-148;  
 Matches 655; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

Qy 1 ATGTGTCACCTCGCAGCTGCGACCCGACCATGACCATCTCGAGGCCCGCCCGGCC 60  
 Db 47 ATGTGTCACCTCGCAGCTGCGACCCGACCATGACCATCTCGAGGCCCGCCCGGCC 106  
 Qy 61 CCTCTCACATCCCGGACCCCGGGGCTCCGGTCTGTAGATCTTACCTTCGACCT 120  
 Db 107 CCTCTCACATCCCGGACCCCGGGGCTCCGGTCTGTAGATCTTACCTTCGACCT 166  
 Qy 121 CTCCTGGAGCGGACCGGCGCCCTGCCGGGGCCCGCCAGCGCTCTCGGGGACCGAAG 180  
 Db 167 CTCCTGGAGCGGACCGGCGCCCTGCCGGGGCCCGCCAGCGCTCTCGGGGACCGAAG 226  
 Qy 181 CCGACCCCGAGGCTCTTACCTCGAGTGGTCCGGGCCAGCTGCCAGTCGAGGAACCG 240  
 Db 227 CCGACCCCGAGGCTCTTACCTCGAGTGGTCCGGGCCAGCTGCCAGTCGAGGAACCG 286  
 Qy 241 AACCCAGCCAAAAGGCTCTCTCTCTGCTGTACCATCGTCTTCTGCGAGATCTTGATG 300  
 Db 287 AACCCAGCCAAAAGGCTCTCTCTCTGCTGTACCATCGTCTTCTGCGAGATCTTGATG 346  
 Qy 301 GCTGAAGAGGTGTGCGGGCGCCCTGCTCCAGAGGACGCCCTTAACCGCGCATCCCTG 360  
 Db 347 GCTGAAGAGGTGTGCGGGCGCCCTGCTCCAGAGGACGCCCTTAACCGCGCATCCCTG 406  
 Qy 361 GGGCCACCCCTGTGTCGCCCGCTCTCGAGGCCCTTAACTGACCTTCGGAGCCCTCGAC 420  
 Db 407 GGGCCACCCCTGTGTCGCCCGCTCTCGAGGCCCTTAACTGACCTTCGGAGCCCTCGAC 466  
 Qy 421 TACGCTCTGGACCTCAGCAGCTTCTCTCCAGCAACCCCGCGCTTCTTAACTGTGACCT 480  
 Db 467 TACGCTCTGGACCTCAGCAGCTTCTCTCCAGCAACCCCGCGCTTCTTAACTGTGACCT 526  
 Qy 481 CCGCAGCTCCCAAAAGAAATCCGAAAACACAAAGAAACACAGCGGTACCTGGTGGC 540  
 Db 527 CCGCAGCTCCCAAAAGAAATCCGAAAACACAAAGAAACACAGCGGTACCTGGTGGC 586  
 Qy 541 CGAGAGCTATCCCAACTGGGACTTCGAGGCAACTTGAACACAGACACTACAGCGGA 600  
 Db 587 CGAGAGCTATCCCAACTGGGACTTCGAGGCAACTTGAACACAGACACTACAGCGGA 646  
 Qy 601 GACGCCACCCGCTCTTGAAGCGGACCGAGGCGCACAGAGACCGAGCGCATAGAGAC 659  
 Db 647 GACGCCA-CCCGTCTTGAGCGGGA-CGAGGCGCACAGAGACC-AGCGGCTATAGAGAC 702

RESULT 3  
 BE621346  
 LOCUS 601493736F1 NIH\_MGC\_70 Homo sapiens cDNA clone IMAGE:3895736 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE621346  
 VERSION BE621346.1 GI:9892286  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 762)  
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNI)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LILNI at:  
 http://image.llnl.gov  
 Plate: LICM674 row: h column: 09  
 High quality sequence stop: 613.  
 Location/Qualifiers  
 1..762  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="3895736"  
 /clone\_lib="NIH\_MGC\_70"  
 /tissue\_type="epithelioid carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site: 1: Nottl;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.1 kb. Library constructed by Life Technologies."

HASH COUNT 155 a 288 c 193 g 126 t

ORIGIN

Query Match 49.2%; Score 604.2; DB 110; Length 762;  
 Best Local Similarity 93.9%; Pred. No. 2.2e-143;  
 Matches 705; Conservative 0; Mismatches 38; Indels 8; Gaps 7;

Qy 2 TGTGTCACTCTCGCAGTGCACCGACCATGACCATCTCGAGGCCCGACCGGCC 61  
 Db 1 TGTGTCACTCTCGCAGTGCACCGACCATGACCATCTCGAGGCCCGACCGGCC 59  
 Qy 62 CCTCACCATCCCGGACCCCGGGGCTCCGGTCTGTAGATCTTACCTTCGACCTC 121  
 Db 60 CCTCACCATCCCGGACCCCGGGGCTCCGGTCTGTAGATCTTACCTTCGACCTC 119  
 Qy 122 TCCCGGAGCCCGACGCGCCCTCGCGGGCCCGCCAGGCCCTCTCGGGGACCGAAGC 181  
 Db 120 TCCCGGAGCCCGACGCGCCCTCGCGGGCCCGCCAGGCCCTCTCGGGGACCGAAGC 179  
 Qy 182 GCAGCCGACAGGTCTCTACCTCGAGTGTCCGCGCCAGCTGCAGTCCAGGAACCGA 241  
 Db 180 GCAGCCGACAGGTCTCTACCTCGAGTGTCCGCGCCAGCTGCAGTCCAGGAACCGA 239  
 Qy 242 ACCCAGCCAAAGGCTTCTCTTCTGTGTCTACCATCGTCTTCTGTGCTTCTTCTG 301  
 Db 240 ACCCAGCCAAAGGCTTCTCTTCTGTGTCTACCATCGTCTTCTGTGCTTCTTCTG 299  
 Qy 302 CTGAAGAGGTGTCCCGGCCCTCGCTCCAGAGGACGCCCTTAAGCCGCATCCCTGG 361  
 Db 300 CTGAAGAGGTGTCCCGGCCCTCGCTCCAGAGGACGCCCTTAAGCCGCATCCCTGG 359  
 Qy 362 CGCCACCCCTGTGTCCCGCTCTCGAGGCCCTTTAATCTGACCTTCGAGGCCCTCGGACT 421  
 Db 360 CGCCACCCCTGTGTCCCGCTCTCGAGGCCCTTTAATCTGACCTTCGAGGCCCTCGGACT 419  
 Qy 422 ACGTCTGGACCTCAGCAGCTTTCCTCCAGCAACACCCGGCGGCTTCTTAACTGTGAC 481  
 Db 420 ACGTCTGGACCTCAGCAGCTTTCCTCCAGCAACACCCGGCGGCTTCTTAACTGTGAC 479  
 Qy 482 CCGCAGTCCCAAAAGAA-TCCGAAAACACAAAGAAACACAGCGGTACCTGGTGGC 540  
 Db 480 CCGCAGTCCCAAAAGAA-TCCGAAAACACAAAGAAACACAGCGGTACCTGGTGGC 539  
 Qy 541 CGAGAGCGGTAT-CCCCAACTGGGACTTCCGAGGCAACTTGAACTCAGAACACTACAGC 597  
 Db 540 CGAGAGCGGTATCCCCAACTGGGACTTCCGAGGCAACTTGAACTCAGAACACTACAGC 599



QY 598 GGAGACCCACCCGGTCTTTCAG-GCGGACCGA-GGCCACACAGACCCGAGC-CCATA 654  
 Db 600 GAGCAGCCACCCGGTCTTTCAGCGCGGACCGAGCGGACAGACAGAGCGGCATA 659  
 QY 655 GAGACCCAGCAGACCCAGCTGGGCTAGGCGCGGTGGGAAGAGAGAGGCTCTTAATTT 714  
 Db 660 GAGACCCAGCAGACAGACAGCTGGGCTAGGCGCGGTGGGAAGAGAGAGGCTCCGCAATTT 719  
 QY 715 ATTCTTATTGCTCTTAATTAATTAATATATAT 745  
 Db 720 ATTACTTATGCTCTTAATTAATTAATAGGTAT 750

RESULT 4  
 BE882130  
 LOCUS 601505227F2 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:3906823 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE882130  
 VERSION BE882130  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLCW703 row: f column: 08  
 High quality sequence stop: 659.

FEATURES  
 source  
 1..916  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="3906823"  
 /clone\_lib="NIH\_MGC\_71"  
 /tissue\_type="leiomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: uterus; Vector: pCMV-Sp10T6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 2.1 kb."  
 BASE COUNT 187 a 327 c 268 g 134 t  
 ORIGIN

Query Match 48.0%; Score 590; DB 137; Length 916;  
 Best Local Similarity 94.7%; Pred. No. 9.4e-140;  
 Matches 676; Conservative 0; Mismatches 30; Indels 8; Gaps 6;

QY 31 ATGACCATCTGCAGCCCGGACCCCGGCGGCTTCCACCATCCGGGACCCCGGGGGGCG 90  
 Db 1 ATGACCATCTGCAGCCCGGACCCCGGCGGCTTCCACCATCCGGGACCCCGGGGGGCG 59  
 QY 91 TCGGTCCTGAGATCTTACCTTCGACCTCTCCGGAGCCCGGAGCGGCGGCTCGCGG 150  
 Db 60 TCGGTCCTGAGATCTTACCTTCGACCTCTCCGGAGCCCGGAGCGGCGGCTCGCGG 119  
 QY 151 CGCCCGAGCGCTCTCCGGGACCCGAAAGCGAGCGCGAGGTTCTTACCTCGAGTG 210  
 Db 120 CGCCCGAGCGCTCTCCGGGACCCGAAAGCGAGCGCGAGGTTCTTACCTCGAGTG 179  
 QY 211 GTCCGGCGCAGCTGCCAGTCGAGGAGACCCAGCCAGCAAGAGCTTCCTCTCTGCTG 270

Db 180 GTCCGGCGCAGTGCACGTTCAGGAACCGAACCCAGCCCAAGAGGCTTCCTTCCTGCTG 239  
 QY 271 CTCACCATCGTCTTCTGCGCAGATCTCTGATGGCTGAAGAGGGTGTGCGCGCGCCCTGCCT 330  
 Db 240 CTCACCATCGTCTTCTGCGCAGATCTCTGATGGCTGAAGAGGGTGTGCGCGCGCCCTGCCT 299  
 QY 331 CCAGAGGAGCGCCCTTAACGCCCGATCCCTTGGGCGCCACCCCTGTGTCTCCCTCTCGAG 390  
 Db 300 CCAGAGGAGCGCCCTTAACGCCCGATCCCTTGGGCGCCACCCCTGTGTCTCCCTCTCGAG 359  
 QY 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGAGCTCAGACTTCCTCTCCAG 450  
 Db 360 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGAGCTCAGACTTCCTCTCCAG 419  
 QY 451 CAACACCGGCGC-CGCTTCTTAATCTGACTTCCTGAGCTCCCGACACTCCCAAAAGAAATCCGAAAA 509  
 Db 420 CAACACCGGCGCCTTCTAATCTGACTTCCTGAGCTCCCGACACTCCCGAAAAAGAAATCCGAAAA 479  
 QY 510 CCACAAAGAAACACACGAGCGGTACCTGGTGGCGGAGAGCTATCCCAACTGGGACTTCGCG 569  
 Db 480 CCACAAAGAAACACACGAGCGGTACCTGGTGGCGGAGAGCTATCCCGAACCTGGGACTTCGCG 539  
 QY 570 AGCCAACTTGAATCTCAGAACTACAGCGGAGACCCAGCCCGTGTCTGAGCGGGGACCG 629  
 Db 540 AGCCAACTTGAATCTCAGAACTACAGCGGAGACCCAGCCCGTGTCTGAGCGGGGACCG 597  
 QY 630 AGGCGCACAGACCGCGGCGGATAGAGACCGGAGGACAGCGGCTGGGCTAGGCGCGG 689  
 Db 598 AGGCGCACAGACCGCGGCGGATAGAGACCGGAGAGC--AGCTGGCGCTAGGCGG 653  
 QY 690 GTGGGAAGGAGAGCGGCTTAATTTATTTCTTATTTGCTTCCTTAATTAATTTAT 743  
 Db 654 GTGGGAAGGAGAGCGGCTTAATTTATTTCTTATTTGCTTCCTTAATTAATTTAT 707

RESULT 5  
 BE273386  
 LOCUS 601143493F1 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:3507191 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE273386  
 VERSION BE273386  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: DCTD/BNP  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCM186 row: j column: 24  
 High quality sequence stop: 579.

FEATURES  
 source  
 1..603  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="3507191"  
 /clone\_lib="NIH\_MGC\_14"  
 /tissue\_type="renal cell adenocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: kidney; Vector: pONH7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

HASH COUNT 112 a 241 c 154 g 96 t

ORIGIN

Query Match 48.0%; Score 589.4; DB 106; Length 603;  
Best Local Similarity 99.7%; Pred. No. 1.3e-139;  
Matches 601; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 35 CCATCTTCGACGCGCCGACCCGCGCCCTCCACCATCCGGACACCGGGGGCTCGG 94  
Db 1 CCATCTTCGACGCGCCGACCCGCGCCCTCCACCATCCGGACACCGGGGGCTCGG 60

QY 95 GTCTGTGAGATCTTACCTTTCGACCTCTCCGGAGCCCGACGCGCCCTGCGGGCGCC 154  
Db 61 GTCTGTGAGATCTTACCTTTCGACCTCTCCGGAGCCCGACGCGCCCTGCGGGCGCC 120

QY 155 CCAGCGCTCTCCGGGCGACCGAAGCGGACGCGGAGGTTCTTACCTTCGAGTGGTCC 214  
Db 121 CCAGCGCTCTCCGGGCGACCGAAGCGGACGCGGAGGTTCTTACCTTCGAGTGGTCC 180

QY 215 GCGCGCAGCTGCGAGTCGAGGAACGACCCAGCCAAAGGCTTCTTCTGCTGCTCA 274  
Db 181 GCGCGCAGCTGCGAGTCGAGGAACGACCCAGCCAAAGGCTTCTTCTGCTGCTCA 240

QY 275 CCATCTCTTCTGCGCAGATCTCGATGGGTGAAGAGGGTGTGCGGCGCCCTGCGTCCAG 334  
Db 241 CCATCTCTTCTGCGCAGATCTCGATGGGTGAAGAGGGTGTGCGGCGCCCTGCGTCCAG 300

QY 335 AGATGCGCCCTTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCC -CGTCTCTGAGGCC 393  
Db 301 AGATGCGCCCTTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCC -CGTCTCTGAGGCC 360

QY 394 TTTAATCTGACTTCGGAGCCCTCGGACTACGCTTGGACCTCAGCACTTTCTCCAGCA 453  
Db 361 TTTAATCTGACTTCGGAGCCCTCGGACTACGCTTGGACCTCAGCACTTTCTCCAGCA 420

QY 454 CACCGCGCGCCCTTCTAATCTGACTTCGCGGACTTCCCGACTTCCGAAACAC 513  
Db 421 CACCGCGCGCCCTTCTAATCTGACTTCGCGGACTTCCCGACTTCCGAAACAC 480

QY 514 AAAGAAACACACGCGGTACCTGGTGGCGGAGAGCTATCCCAACTGGGACTTCCGAGGC 573  
Db 481 AAAGAAACACACGCGGTACCTGGTGGCGGAGAGCTATCCCAACTGGGACTTCCGAGGC 540

QY 574 AACTTGAACCTCAGAACACTACAGCGGAGACGCCACCCGGTGTGAGGCGGAGCCGAGGC 633  
Db 541 AACTTGAACCTCAGAACACTACAGCGGAGACGCCACCCGGTGTGAGGCGGAGCCGAGGC 600

QY 634 GCA 636  
Db 601 GCA 603

RESULT 6  
HE563592 810 bp mRNA EST 15-AUG-2000  
LOCUS 601:4706F1 NIH\_MGC\_39 Homo sapiens cDNA clone IMAGE:368825 5',  
DEFINITION mRNA sequence.  
ACCESSION BE563592  
VERSION BE563592.1 GI:9807312  
KEYWORDS EST.  
SOURCE Human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 810)  
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

TITLE  
JOURNAL  
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (I.M.N.I.)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LCM382 row: k column: 02  
High quality sequence stop: 692.  
Location/Qualifiers  
1. 810  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:368825"  
/clone\_lib="NIH\_MGC\_39"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed  
by Ling Hong in the laboratory of Gerald M. Rubin  
University of California, Berkeley) using ZAP-cDNA  
synthesis kit (Stratagene) and Superscript II RT (Life  
Technologies)."

FEATURES  
source

BASE COUNT 176 a 284 c 205 g 145 t  
ORIGIN

Query Match 47.9%; Score 588; DB 109; Length 810;  
Best Local Similarity 92.1%; Pred. No. 3e-139;  
Matches 753; Conservative 0; Mismatches 50; Indels 15; Gaps 12;

QY 1 ATGTGTCACTTCGACGTGCGACCCCGGCGGTCTCGAGATCTTCGACCTTCGACCCCT 120  
Db 1 ATGTGTCACTTCGACGTGCGACCCCGGCGGTCTCGAGATCTTCGACCTTCGACCCCT 60

QY 61 CCTTCCACCATCCCGGACCCCGGCGGTCTCGAGATCTTCGACCTTCGACCCCT 120  
Db 58 CCTTCCACCATCCCGGACCCCGGCGGTCTCGAGATCTTCGACCTTCGACCCCT 116

QY 121 CTCCGGAGCGCGGCGCGCCCTTCCCTCCAGAGGAGCGCCCTTAACGCCGATCCC-T 359  
Db 117 CTCCGGAGCGCGGCGCGCCCTTCCCTCCAGAGGAGCGCCCTTAACGCCGATCCC-T 356

QY 181 CGCAGCGCGAGGGTTCCTTACCTTCGAGTGTCTCCGGCGGCGAGTGCAGTGCAGGAACCG 240  
Db 177 CGCAGCGCGAGGGTTCCTTACCTTCGAGTGTCTCCGGCGGCGAGTGCAGTGCAGGAACCG 236

QY 241 AACCCAGCCAAAAGGCTTCTTCTTCTGCTGCTCACCATCGTCTTCTGCGAGATCCGTGATG 300  
Db 237 AACCCAGCCAAAAGGCTTCTTCTTCTGCTGCTCACCATCGTCTTCTGCGAGATCCGTGATG 296

QY 301 GCTGAAGAGGGTGTGCGGCGCCCTTCCCTCCAGAGGAGCGCCCTTAACGCCGATCCC-T 359  
Db 297 GCTGAAGAGGGTGTGCGGCGCCCTTCCCTCCAGAGGAGCGCCCTTAACGCCGATCCC-T 356

QY 360 GCGCGCCACCCCTGTGTCCCGCGTCTCGAGCCCTTTAATCTGACTTCGAGAGCCCTCGGA 419  
Db 357 GCGCGCCACCCCTGTGTCCCGCGTCTCGAGCCCTTTAATCTGACTTCGAGAGCCCTCGGA 416

QY 420 CTACGCTCTGGACCTCAGCACTTTCCCTCCAGCAACACCGCGCCCTTCTTAAGTGTGACT 479  
Db 417 CTACGCTCTGGACCTCAGCACTTTCCCTCCAGCAACACCGCGGAC--GCTTCTTAAGTGTGACT 474

QY 480 CCCCAGCTCCCAAAAAGAAATCCGAAAACACCAAAAGAACACACGAGCGTACCTGGTGC 539  
Db 475 CCCCAGCTCCCAAAAAGAAATCCGAAAACACCAAAAGAACACACGAGCGTACCTGGTGC 534

QY 540 GCGAGAGCGTATCCCACTGG-GACTCTCGAGGCAACTTGAACCTCAGAAC-ACTACAGC 597  
 DB 535 GCGAGAGCGTATCCCACTGGAGACTTCGAGGCAACTTGAACCTCAGAACACTACAGC 594  
 QY 598 GGAGAGCGCCACACG-CGGTCTGTAGAGCGGAGCGAGGCGCA-CAGAGACCGAGG-CGCAT 553  
 DB 595 GGAGAGCGCCACACCTGGTGTCTGTAGAGCGGAGCGAGGCGG777CAAAGACCGAGGCGCAT 654  
 QY 654 AGAGACCGAGGACACCGCCAGCTGGGGCTAGGCCCGTGGGAGGAGGACCGCTCGTTAAAT 713  
 DB 655 AGAGACCGAGGACACG-CAGATAGGCGCTAGGCCCGTGGGAGGAGGAGCGTCCGCAAT 713  
 QY 714 TATTCTTATGCTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 773  
 DB 714 TAATATCTTATCTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 772  
 QY 774 GAGATGTGACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 811  
 DB 773 GACAAATTGTACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 810

RESULT 7  
 BE910077 585 bp mRNA EST 29-SEP-2000  
 LOCUS 601502143F1 NIH\_MGC\_70 Homo sapiens cDNA clone IMAGE:3903754 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE910077  
 VERSION BE910077.1 GI:10406309  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 585)  
 NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLCM695 row: f column: 11  
 High quality sequence stop: 585.  
 FEATURES  
 source  
 1..585  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3903754"  
 /clone\_lib="NIH\_MGC\_70"  
 /tissue\_type="epithelioid carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.1 kb. Library constructed by Life  
 Technologies."  
 BASE COUNT 131 a 119 c 188 g 147 t  
 ORIGIN

Query Match  
 Best Local Similarity 47.6% Score 585; DB 137; Length 585;  
 Matches 595; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 544 GACCGTATCCCACTGGGACTTCGAGGCAACTTGAACCTCAGAACACTACAGCGGAGAC 603  
 DB 1 GACCGTATCCCACTGGGACTTCGAGGCAACTTGAACCTCAGAACACTACAGCGGAGAC 60

QY 604 GCCACCCGCTGTAGAGCGGAGCAGGCGCACAGAGACCGAGGCGCATAGAGACCGAG 663  
 DB 61 GCCACCCGCTGTAGAGCGGAGCAGCAGCGCACAGAGACCGAGGCGCATAGAGACCGAG 120  
 QY 664 GCACAGCCAGCTGGGGCTAGGCCCGTGGGAGGAGAGGCTCGTTAATTTATTTCTTAT 723  
 DB 121 GCACAGCCAGCTGGGGCTAGGCCCGTGGGAGGAGAGGCTCGTTAATTTATTTCTTAT 180  
 QY 724 TCTCTCTAAATTAATATTTATATGTTATTTATGTTACGTCCTCCTAGGATGTTGA 783  
 DB 181 TCTCTCTAAATTAATATTTATATGTTATTTATGTTACGTCCTCCTAGGATGTTGA 240  
 QY 784 CGTAATATTTATTTAACTTATGCAAGGCTGTGAGATGTTCCCTCTGCTAATGCAGG 843  
 DB 241 CGTAATATTTATTTAACTTATGCAAGGCTGTGAGATGTTCCCTCTGCTAATGCAGG 300  
 QY 844 TCTCTTGGTATTTATTTAGCTTTTGTGGACTTGTGGAAGCAGGACACCTCGGAACCTCGGC 903  
 DB 301 TCTCTTGGTATTTATTTAGCTTTTGTGGACTTGTGGAAGCAGGACACCTCGGAACCTCGGC 360  
 QY 904 AAAGTAGGAGAAATGGGAGGACTCGGCTGGGGGAGGAGCTCCCGGCTGGGATCAAG 963  
 DB 361 AAAGTAGGAGAAATGGGAGGACTCGGCTGGGGGAGGAGACGTCGCGGCTGGGATCAAG 420  
 QY 964 TCTGTGGTGGGTCGTAAGTTTAGGAGGTGACTGTCATCTCCAGCATCTCAACTCCGCTCT 1023  
 DB 421 TCTGTGGTGGGTCGTAAGTTTAGGAGGTGACTGTCATCTCCAGCATCTCAACTCCGCTCT 480  
 QY 1024 GTCTACTGTGTGAGACTTCGGGGACCATTAGGAATGAGATCCGTGAGATCCCTTCCATCT 1083  
 DB 481 GTCTACTGTGTGAGACTTCGGGGACCATTAGGAATGAGATCCGTGAGATCCCTTCCATCT 540  
 QY 1084 TCTTGAAGTCGCTTTAGGCTGGCTCGGAGGATAGAGGTTGGGG 1128  
 DB 541 TCTTGAAGTCGCTTTAGGCTGGCTCGGAGGATAGAGGTTGGGG 585

RESULT 8  
 AW340002/c 586 bp mRNA EST 31-JAN-2000  
 LOCUS hc91d03.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
 DEFINITION IMAGE:2907365 3', mRNA sequence.  
 ACCESSION AW340002  
 VERSION AW340002.1 GI:6836628  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 586)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 471.  
 FEATURES  
 Location/Qualifiers  
 1..586  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2907365"  
 /clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: p773p-pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NBH19W, testis NIH, and B-cell

NCI\_CGAP\_CCB1} were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 155 a 187 c 110 g 134 t

Query Match 47.3%; Score 581.2; DB 89; Length 586;  
Best Local Similarity 99.5%; Pred. No. 1.5e-137;  
Matches 583; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 621 GCGGGACCGGCGGCACAGAGCCGAGCGGCATAGAGCCGAGGCACGCCAGCTGGGG 680  
DB 586 GCGGGACCGGCGGCACAGAGCCGAGCGGCATAGAGCCGAGGCACGCCAGCTGGGG 527  
QY 681 CTAGGCGCGGTGGGAAGGAGAGCGTCTTAATTTATTTCTTATCTCTCTAATTAAT 740  
DB 526 CTAGGCGCGGTGGGAAGGAGAGCGTCTTAATTTATTTCTTATCTCTCTAATTAAT 467  
QY 741 TATATCTATTTATGTACGCTCCTAGGTGATGGAGATGTACGTAATATTTATTTAA 800  
DB 466 TATATCTATTTATGTACGCTCCTAGGTGATGGAGATGTACGTAATATTTATTTAA 407  
QY 801 CTATCCAGAGGTGTGAGATGTCTCTGTGTAAATGCAGCTCTCTTGGTATTATG 860  
DB 406 CTATCCAGAGGTGTGAGATGTCTCTGTGTAAATGCAGCTCTCTTGGTATTATG 347  
QY 861 AGCTTTGTGGGACTGTGTGAAGCAGACACCTGGACCTGCGCAAGTAGGAGAGAAAT 920  
DB 346 AGCTTTGTGGGACTGTGTGAAGCAGACACCTGGACCTGCGCAAGTAGGAGAGAAAT 287  
QY 921 GGGGAGCACTCCGGTGGGGAGGAGCCTCCCGGCTGGGATGAAGTCTGCGTGGGCGTA 980  
DB 286 GGGGAGCACTCCGGTGGGGAGGAGCCTCCCGGCTGGGATGAAGTCTGCGTGGGCGTA 227  
QY 981 AGTTTAGGAGGCTAGTCATCTCCAGCATCTCACTCCGCTGTCTACTGTGTGAGAT 1040  
DB 226 AGTTTAGGAGGCTAGTCATCTCCAGCATCTCACTCCGCTGTCTACTGTGTGAGAT 167  
QY 1041 TCGGCGGACCATTAGGAATGAGATCCGTGAGATCTTCCATCTCTTGAAGTCGCTTTA 1100  
DB 166 TCGGCGGACCATTAGGAATGAGATCCGTGAGATCTTCCATCTCTTGAAGTCGCTTTA 107  
QY 1101 GGGTGGCTCCGAGGTAGAGGTTGGGGTTCGTGGGCTGTACCGGAGGCACTGTCCGAG 1160  
DB 106 GGGTGGCTCCGAGGTAGAGGTTGGGGTTCGTGGGCTGTACCGGAGGCACTGTCCGAG 47  
QY 1161 GCGCTAGTATGTTCTGTGAACAAATAAATTTGATTTACTGTCAA 1206  
DB 46 GCGCTAGTATGTTCTGTGAACAAATAAATTTGATTTACTGTCAA 1

RESULT 9  
LOCUS AW662348/c 598 bp mRNA EST 06-APR-2000  
DEFINITION h125c07.x1 NCI\_CGAP\_Col4 Homo sapiens cDNA clone IMAGE:2973324 3', mRNA sequence.  
ACCESSION AW662348  
VERSION AW662348.1 GI:7454886  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 598)  
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CCAP),  
TITLE Tumor Gene Index

# JOURNAL COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA library Preparation: Life Technologies, Inc.  
cDNA library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
image.llnl.gov/image/html/iresources.shtml

Seq primer: -400P from Gibco  
High quality sequence stop: 392.

## FEATURES

Location/Qualifiers  
1..598  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2973324"  
/clone\_lib="NCI\_CGAP\_Col4"  
/issue\_type="moderately-differentially-adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site:1: SalI;  
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.7 kb. Life Technologies catalog #:  
11531-019"

BASE COUNT 163 a 187 c 116 g 132 t

## ORIGIN

Query Match 47.2%; Score 579.6; DB 93; Length 598;  
Best Local Similarity 98.5%; Pred. No. 3.9e-137;  
Matches 585; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 611 GGTGCTTGAGCGGACCGGCGGCACAGAGCCGAGGCATAGAGCCGAGGCACGCCAG 670  
DB 598 GGTGCTTGAGCGGACCGGCGGCACAGAGCCGAGGCATAGAGCCGAGGCACGCCAG 539  
QY 671 CAGCTGGGGCTAGGCGGCGGTGGGAAGGAGAGCGTCTTAATTTATTTCTTATCTCTCT 730  
DB 538 CAGCTGGGGCTAGGCGGCGGTGGGAAGGAGAGCGTCTTAATTTATTTCTTATCTCTCT 479  
QY 731 AATTAATATTTATATGATTTATGACCTCTCTAGTGATGGAGATGTCTACCTAATA 790  
DB 478 AATTAATATTTATATGATTTATGACCTCTCTAGTGATGGAGATGTCTACCTAATA 419  
QY 791 TTTATTTAACTTATGCAAGGTGTGAGATGTCTCTCTGTAAATGCAAGTCTCTTTG 850  
DB 418 TTTATTTAACTTATGCAAGGTGTGAGATGTCTCTCTGTAAATGCAAGTCTCTTTG 359  
QY 851 GTATTTATGAGCTTTGTGGGACTGGTGAAGAGGACACCTGGAAATTCGGGCAAAATAG 910  
DB 358 GTATTTATGAGCTTTGTGGGACTGGTGAAGAGGACACCTGGAAATTCGGGCAAAATAG 299  
QY 911 GAGAAGAAATCGGAGGACTCGGGTGGGGAGGAGGACCTCCCGGCTGGGATGAAGTCTGGTG 970  
DB 298 GAGAAGAAATCGGAGGACTCGGGTGGGGAGGAGGACCTCCCGGCTGGGATGAAGTCTGGTG 239  
QY 971 GTGGGTCTTAAGTTTAGGAGGTGACTGCTCCAGCATCTCAACTCCGCTCTGCTACT 1030  
DB 238 GTGGGTCTTAAGTTTAGGAGGTGACTGCTCCAGCATCTCAACTCCGCTCTGCTACT 179  
QY 1031 GTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCTCTTCCATCTTCTTGA 1090  
DB 178 GTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCTCTTCCATCTTCTTGA 119  
QY 1091 GTCCGCTTTAGGCTGGCTGCCAGGTAGAGGTTGGGGTTCGGTGGCTGTGCACGAGCGA 1150  
DB 118 GTCCGCTTTAGGCTGGCTGCCAGGTAGAGGTTGGGGTTCGGTGGCTGTGCACGAGCGA 59  
QY 1151 CTGTGAGAGTCCGCTAGTATGTTCTGTGAACACAAATAAATTTGATTTACTGTCT 1204  
DB 58 CTGTGAGAGTCCGCTAGTATGTTCTGTGAACACAAATAAATTTGATTTACTGTCT 5

RESULT 10  
 BE621600/c  
 LOCUS  
 DEFINITION  
 BE621600 614 bp mRNA EST 24-AUG-2000  
 601493736T1 NIH\_MGC\_70 Homo sapiens cDNA clone IMAGE:3895736 3',  
 mRNA sequence.  
 BE621600  
 BE621600.1 GI:9892540  
 EST.  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 614)  
 NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue procurement: ATCC  
 cDNA Library preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LCM674 row: h column: 09  
 High quality sequence start: 12  
 High quality sequence stop: 613.  
 High quality sequence stop: 613.

#### FEATURES

source  
 1..614  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3895736"  
 /clone\_lib="NIH\_MGC\_70"  
 /tissue\_type="epitheloid carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site:1: NotI;  
 Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.1 kb. Library constructed by Life  
 Technologies."

BASE COUNT 158 a 198 c 123 g 135 t  
 ORIGIN

Query Match 46.2%; Score 567.2; DB 110; Length 614;  
 Best Local Similarity 98.2%; Pred. No. 5.7e-134;  
 Matches 595; Conservative 0; Mismatches 8; Indels 3; Gaps 2;

QY 582 CTCAGAACACTACAGCGGAGACCCACCGGTGCTTGAGCGGCGACCGGCGACAG 641  
 |||||||  
 DB 614 CTCAGAACACTACAGCGGAGACCCACCGGTGCTTGAGCGGCGACCGGCGACAG 555  
 |||||||  
 QY 642 ACCGAGCGCATAGACCGAGCGACCGACCGACCTGGGCTAGGCCCGGTGGGAAGAGA 701  
 |||||||  
 DB 554 ACCGAGCGCATAGACCGAGCGACCGACCGACCGACCTGGGCTAGGCCCGGTGGGAAGAGA 495  
 |||||||  
 QY 702 GCCTCGTTAA--TTTATTCTTATTGCTCTCTAATTAATATTTATATGTTATGTTACGT 759  
 |||||||  
 DB 494 GCCTCGTTAATGTTATGTTCTTATTGCTCTCTAATTAATATTTATGTTATGTTACGT 435  
 |||||||  
 QY 760 CTTCTTAGCTAGTACATGTGTACCTAATATTTATTTAACTATGCAAGGCGTGTGAGA 819  
 |||||||  
 DB 434 CTTCTTAGCTAGTACATGTGTACCTAATATTTATTTAACTATGCAAGGCGTGTGAGA 375  
 |||||||  
 QY 820 TGTTCCTCTGCTGTAATGTCAGGCTCTTGTGTTATTTATGACTTTGTGGGACTGGTGG 879  
 |||||||  
 DB 374 TGTTCCTCTGCTGTAATGTCAGGCTCTTGTGTTATTTATGACTTTGTGGGACTGGTGG 315  
 |||||||  
 QY 880 AAGCAGGACACCTGGAACTGCGGCAAGTAGGAGAGAATGGGAGGA-CTCGGCTGGG 938  
 |||||||

DB 314 AAGCAGGACACCTGGAACTCGCGCAAGTAGGAGAGAATAATGGGAGGATCTTCGGGTGGG 255  
 QY 939 GGAGGACGTCCTCCGCTGGGATGAAGTCTGCTGCTGGTCTCTAAGTCTTAGGAGGTGACTCC 998  
 |||||||  
 DB 254 GGAGGACGTCCTCCGCTGGGATGAAGTCTGCTGCTGGTCTCTAAGTCTTAGGAGGTGACTCC 195  
 |||||||  
 QY 999 ATCCTCCAGCATCTCAACTCCGCTGCTCTACTGTGTGAGACTTCGGCGGACCATTAGGAA 1058  
 |||||||  
 DB 194 ATCCTCCAGCATCTCAACTCCGCTGCTCTACTGTGTGAGACTTCGGCGGACCATTAGGAA 135  
 |||||||  
 QY 1059 TGAGATCCGTGAGATCCTTCCATCTCTTGAAGTCGCCCTTTAGGCTGGCGGAGGTAGA 1118  
 |||||||  
 DB 134 TGAGATCCGTGAGATCCTTCCATCTCTTGAAGTCGCCCTTTAGGCTGGCGGAGGTAGA 75  
 |||||||  
 QY 1119 GGGTTGGGGTTCGTTGGGCTGTCCAGGAGCGACTTCGAGATCCCTACTATCTTCTCTG 1178  
 |||||||  
 DB 74 GGGTTGGGGTTCGTTGGGCTGTCCAGGAGCGACTTCGAGATCCCTACTATCTTCTCTGGA 15  
 |||||||  
 QY 1179 AACACA 1184  
 |||||  
 DB 14 CACAAA 9

RESULT 11  
 AW956285  
 LOCUS  
 DEFINITION  
 AW956285 588 bp mRNA EST 01-JUN-2000  
 EST368355 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.  
 ACCESSION  
 VERSION  
 AW956285.1 GI:8145968  
 EST.  
 KEYWORDS  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 588)  
 Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holl,  
 I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J., and  
 Quackenbush, J.  
 Assessment of gene expression patterns in a model of colon tumor  
 metastasis using a 19,200 element cDNA microarray  
 Unpublished (2000)  
 Contact: John Quackenbush  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 3528  
 Fax: 301 838 0208  
 Email: johnq@tigr.org  
 Plate: 90  
 Seq primer: Reverse.  
 Location/Qualifiers  
 1..588  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="MAGE resequences, MAGD"  
 /note="Vector: pBluescriptSkm"

BASE COUNT 135 a 117 c 187 g 149 t  
 ORIGIN

Query Match 45.7%; Score 561.6; DB 97; Length 588;  
 Best Local Similarity 98.3%; Pred. No. 1.5e-132;  
 Matches 578; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 556 AACTGGGACTTCGAGGCAACTTGAACCTCAGAACACTACAGCGGAGACCCACCGGTCG 615  
 |||||||  
 DB 1 AACTGGGACTTCGAGGCAACTTGAACCTCAGAACACTACAGCGGAGACCCACCGGTCG 60  
 |||||||  
 QY 616 TTGAGCGGAGCCGAGCGGCACAGACCGAGCGGCATAGAGACCGGAGGACAGCCAC 675  
 |||||||  
 DB 61 TTGAGCGGAGCCGAGCGGCACAGACCGAGCGGCATAGAGACCGGAGGACAGCCAC 120  
 |||||||  
 QY 676 TGGGGCTAGGCCCGCTGGGAAGGAGAGCGTCTGTTATTTATTTCTTATTTCTCTAATTA 735  
 |||||||

```

Db 121 TCGGCTAGCCCGCTCGGAAGGAGACGCTGCTTAATTTATTTCTTATTTCTCTCTTAATTA 180
Qy 736 ATATTATATGATTTATATGATACGCTCTCTAGGTGATGGAGATGTGTACGTATATATTTAT 795
Db 181 ATATTATATGATTTATATGATACGCTCTCTAGGTGATGGAGATGTGTACGTATATATTTAT 240
Qy 796 TTTTAACTTATTCACACCGCTGTGAGATGTTCCCTCTCTGCTGTAAATTCACGCTCTCTGCTPAAT 855
Db 241 TTTTAACTTATTCACACCGCTGTGAGATGTTCCCTCTCTGCTGTAAATTCACGCTCTCTGCTPAAT 300
Qy 856 TATTGACCTTTTGGGGACTGGTGAAGCAGGACACCTTGGAACTCGGCAAGTAGGAGAA 915
Db 301 TATTGACCTTTTGGGGACTGGTGAAGCAGGACACCTTGGAACTCGGCAAGTAGGAGAA 360
Qy 916 GAAATGGGAGGACTCGGGTGGGGAGGAGCTCCCGCTGGGATGAAGTCTGGTGGGG 975
Db 361 GAAATGGGAGGACTCGGGTGGGGAGGAGCTCCCGCTGGGATGAAGTCTGGTGGGG 420
Qy 976 TCGTAAGTTTGGAGCTGACTGCTCCAGCATCTCAACTCCGCTGCTACTGTGTC 1035
Db 421 TCGTAAGTTTGGAGGCTGACTGCTCCAGCATCTCAACTCCGCTGCTACTGTGTC 480
Qy 1036 AGACTTCGGGGACCATTAGAATGAGATCCGTGGAGATCTTCCATCTTTTGAAGTCGC 1095
Db 481 AGACTTCGGGGACCATTAGAATGAGATCCGTGGAGATCTTCCATCTTTTGAAGTCGC 540
Qy 1096 CTTTAGGCTGGCTGCCAGGT-AGAGGGTTGGGGGTTGGTGGGCTGTCA 1142
Db 541 CTTTAGGCTGGCTGCCAGGT-AGAGGGTTGGGGGTTGGTGGGCTGTAA 588

RESULT 12
AV690660 617 bp mRNA EST 25-SEP-2000
LOCUS AV690660 GK Homo sapiens cDNA clone GKBA04 5', mRNA sequence.
DEFINITION AV690660
ACCESSION AV690660.1 GI:10292523
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
Wu, T., Qian, H., Huang, Q., Kang, B., Gao, X., Xu, Z., Xiao, H.,
Xu, X., Li, N., Peng, Y., Liu, F., Qu, J., Song, H., Cheng, Z., Zeng, L.,
Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang,
Y., Gu, Y., Chen, Z. and Han, Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
Source Location/Qualifiers
1..617
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKBA04"
/clone.lib="GKC"
/lisue.type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 131 a 182 c 164 g 139 t 1 others
ORIGIN

Query Match 45.4% Score 557 DB 38 Length 617;

```

```

Best local Similarity 95.5%; Pred. No. 2.3e-131;
Matches 596; Conservative 0; Mismatches 21; Indels 7; Gaps 2;

Qy 277 ATCGTCTTCTGCGCAGATCCTGATGGCTGAAGAGGGTGTGCCGGGCGCCCTGCTCCACAGAG 336
Db 1 ATCGTCTTCTGCGCAGATCCTGATGGCTGAAGAGGGTGTGCCGGGCGCCCTGCTCCACAGAG 60
Qy 337 GAGCCCTTAAACCCCGCATCCCTTGGGGCCGACCCCTGCTGCCCGCTCCCTCGAGCCCTTT 396
Db 61 GAGCCCTTAAACCCCGCATCCCTTGGGGCCGACCCCTGCTGCCCGCTCCCTCGAGCCCTTT 120
Qy 397 AATCTGACATTCGAGAGCCCTCGGACTAGGCTGTGGACCTCAGCACTTTCCCTCCAGCAACAC 456
Db 121 AATCTGACATTCGAGAGCCCTCGGACTAGGCTGTGGACCTCAGCACTTTCCCTCCAGCAACAC 180
Qy 457 CCGCCCGCTTCTAACTGTGACTCCCGGACCTCCCAAAAGAAATCCGAAACACACAA 516
Db 181 CCGCCCGCTTCTAACTGTGACTCCCGGACCTCCCAAAAGAAATCCG---ACCACAA 236
Qy 517 GAAACACCGAGGCTACCTGGTGGCGGAGAGCGGTATCCCAACTGGGACTTCCGAGGCAAC 576
Db 237 GAAACACCGAGGCTACCTGGTGGCGGAGAGCGGTATCCCAACTGGGACTTCCGAGGCAAC 296
Qy 577 TTCAACTCAGAACACTACAGCGGAGAGCGCACCCGGTGTCTTGAGCGGGAGCGGCGCA 636
Db 297 TTCAACTCAGAACACTACAGCGGAGAGCGCACCCGGTGTCTTGAGCGGGAGCGGCGCA 356
Qy 637 CAGAGACCGAGGCGCATAGAGACCGGAGCGGACCGCTGGGCTAGGCGCGGTGGGA 696
Db 357 CAGAGACCGAGGCGCATAGAGACCGGAGCGGACCGCTGGGCTAGGCGCGGTGGGA 416
Qy 697 GGAGAGCGCTGTAATTTATTTCTTATTTCTCTCTAATTAATATTTATATGTA 756
Db 417 GGAGAGCGCTGTAATTTATTTCTTATTTCTCTCTAATTAATATTTATGTA 476
Qy 757 CGTCTCTCTAGGTGATGGAGATGTGTACGTATATTTTAACTTATGCAAGGCTGC 816
Db 477 CGTCTCTCTAGGTGATGGAGATGTGTACGTATATTTTAACTTATGCAAGGCTGC 536
Qy 817 AGATGTTTCCCTCTGCTGTAATGTCAGGCTCTCTTGGTATTTATTTAGCTTTGTGGACTGG 876
Db 537 AGATGTTTCCCTCTGCTGTAATGTCAGGCTCTCTTGGTATTTATTTAGCTTTGTGGACTGG 593
Qy 877 TCGAAGCAGGACACCTTGGAACTGC 900
Db 594 GGGAAAGCAGCGCCCTGGCTGCGGC 617

RESULT 13
AW263252 548 bp mRNA EST 28-DEC-1999
LOCUS xn79g02.x1 Soares_NFL_T_GBC.S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2700722 3', mRNA sequence.
ACCESSION AW263252
VERSION AW263252.1 GI:6640068
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 548)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 466.
Location/Qualifiers

```

```

source
1. 548
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2700722"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT730-Pac (Pharmacia) with
a modified polylinker; Site_1: Not 1; Site_2: Eco R1;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCI-CGAP-GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "
BASE COUNT 157 a 171 c 98 g 122 t
ORIGIN

Query Match 44.4%; Score 544.8; DB 88; Length 548;
Best Local Similarity 99.6%; Pred. No. 2.8e-128;
Matches 546; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 660 CGAGGCACAGCCACGCTGGGCTAGCGCGGTGGGAAGGAGAGCGTCTTAATTTATTTTC 719
Db 548 CGAGGCACAGCCACGCTGGGCTAGCGCGGTGGGAAGGAGAGCGTCTTAATTTATTTTC 489

QY 720 TTATTTCCTCCTAATTAATTTATATATATATATATATATATATATATATATATAT 779
Db 488 TTATTTCCTCCTAATTAATTTATATATATATATATATATATATATATATATATAT 429

QY 780 TTATAGCTATATTTTATTTTAACTTATGCAAGGCTGTGAGATGTTCCCTCTGCTGTAATG 839
Db 428 TTATAGCTATATTTTATTTTAACTTATGCAAGGCTGTGAGATGTTCCCTCTGCTGTAATG 369

QY 840 CAGGCTCTTGCTATTTTATTTAGCTTTTGTGGACTGGTGGAGCAGGACACCTGGAAGCTG 899
Db 368 CAGGCTCTTGCTATTTTATTTAGCTTTTGTGGACTGGTGGAGCAGGACACCTGGAAGCTG 309

QY 900 CGCCAAAGTACGAGAAATATGGGAGGACTCGCGTGGGGAGGAGACGTCCTGGCTGGAT 959
Db 308 CGCCAAAGTACGAGAAATATGGGAGGACTCGCGTGGGGAGGAGACGTCCTGGCTGGAT 249

QY 960 GAAGTCTGCTGTGGTGTAGTTTATGAGGTGACTGCTCCTCCAGCATCTCAACTCC 1019
Db 248 GAAGTCTGCTGTGGTGTAGTTTATGAGGTGACTGCTCCTCCAGCATCTCAACTCC 189

QY 1020 GTCTGTCTACTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCTTCC 1079
Db 188 GTCTGTCTACTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCTTCC 129

QY 1080 ATCTTTTGAAGTCCCTTTAGGGTGGCTGGCAGGTAGAGGTTGGGGTTGGTGGGCTG 1139
Db 128 ATCTTTTGAAGTCCCTTTAGGGTGGCTGGCAGGTAGAGGTTGGGGTTGGTGGGCTG 69

QY 1140 TCACGAGGACATGTGAGATCCCTAGTATGTTCTGTGAACACAAATAAATTTATTTA 1199
Db 68 TCACGAGGACATGTGAGATCCCTAGTATGTTCTGTGAACACAAATAAATTTATTTA 9

QY 1200 CTGTCAAA 1207
Db 8 CTGTCAAA 1

RESULT 14
BE878459
LOCUS
DEFINITION 601487866F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3890248 5',
mRNA sequence
ACCESSION BE878459

```

```

VERSION BE878459.1 GI:10327235
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1070)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTMD/BTP/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MCC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM660 row: c column: 17
High quality sequence stop: 645.
FEATURES
source
1. 1070
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3890248"
/clone_lib="NIH_MGC_69"
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: Not 1;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 276 a 324 c 302 g 168 t
ORIGIN

Query Match 44.0%; Score 540.8; DB 137; Length 1070;
Best Local Similarity 95.6%; Pred. No. 3.3e-127;
Matches 675; Conservative 0; Mismatches 17; Indels 14; Gaps 11;

QY 151 CGCCCCAGCGCTCTCGCGGCGACCGAAAGCGCAGCGG-CAGGGTTCTCTACCCCTCGA-G 208
Db 1 CGCCCCAGCGCTCTCGCGGCGACCGAAAGCGCAGCGGTTCTCTACCCCTCGATG 60

QY 209 TGGTCCGCGGCCAG-CTGCCAGTCGAGNACCGAACCCAGCCAA-AAGGCTTCTCTTCT 266
Db 61 TGGTCCGCGGCCAGTCTGCCAGTCGAGNACCGAACCCAGCCAAAGGCTTCTCTTCT 120

QY 267 GCTGC-TCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGGCGCCC 325
Db 121 GCTGCATCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGGCGCCC 180

QY 326 TGCTCCAGAGGAC-GCCCTTAACGCCGCATCCCTTGGC-GCCACCCCTGTGTCCCGCG 383
Db 181 TGCTCCAGAGGACTGCCCTTAACGCCGCATCCCTTGGCTGGCCACCCCTGTGTCCCGCT 240

QY 384 CCTCAGGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTGGACCTCAGCAGCTTT 443
Db 241 CCTCAGGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTGGACCTTTCAGCAGCTTT 300

QY 444 CCTCCAGCAACACCCGCGGCGCTTCTAATCTGACTTCCCGCAGCTCCCGCAAAAGAAATTC 503
Db 301 CCTCCAGCAACACCCGCGGCGCTTCTAATCTGACTTCCCGCAGCTCCCGCAAAAGAAATTC 360

QY 504 GAAACACCAAGAAACACACAGGCGTACCTGTCGCCAGAGCGTATCCCAACTGGGA 563
Db 361 GAAACACCAAGAAACACACAGGCGTACCTGTCGCCAGAGCGTATCCCAACTGGGA 420

QY 564 CTTCCGAGGCACTTGAAGTACACTAGACGAGGAGACGACCCCGGTGCTTGAAGGCG 623
Db 421 CTTCCGAGGCACTTGAAGTACACTAGACGAGGAGACGACCCCGGTGCTTGAAGGCG 480

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: January 30, 2001, 22:04:48 : Search time 2136.76 Seconds  
(without alignments)  
509.900 Million cell updates/sec

Title: US-08-799-910-9\_COPY\_211\_468  
Perfect score: 258  
Sequence: 1 GTCCGGCCAGTCCAGT.....AGCAACACCGCGCGCTTC 258

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 12325308 seqs, 211498085 residues  
Total number of hits satisfying chosen parameters: 24650616

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: /cgn2\_6/ptodata/2/pna/PCTUS\_COMB.seq.\*
  - 2: /cgn2\_6/ptodata/2/pna/US06\_COMB.seq.\*
  - 3: /cgn2\_6/ptodata/2/pna/US07\_COMB.seq.\*
  - 4: /cgn2\_6/ptodata/2/pna/US08\_COMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/pna/US081\_COMB.seq.\*
  - 6: /cgn2\_6/ptodata/2/pna/US082\_COMB.seq.\*
  - 7: /cgn2\_6/ptodata/2/pna/US083\_COMB.seq.\*
  - 8: /cgn2\_6/ptodata/2/pna/US084\_COMB.seq.\*
  - 9: /cgn2\_6/ptodata/2/pna/US085\_COMB.seq.\*
  - 10: /cgn2\_6/ptodata/2/pna/US086\_COMB.seq.\*
  - 11: /cgn2\_6/ptodata/2/pna/US087\_COMB.seq.\*
  - 12: /cgn2\_6/ptodata/2/pna/US088\_COMB.seq.\*
  - 13: /cgn2\_6/ptodata/2/pna/US089\_COMB.seq.\*
  - 14: /cgn2\_6/ptodata/2/pna/US090\_COMB.seq.\*
  - 15: /cgn2\_6/ptodata/2/pna/US091\_COMB.seq.\*
  - 16: /cgn2\_6/ptodata/2/pna/US092\_COMB.seq.\*
  - 17: /cgn2\_6/ptodata/2/pna/US093\_COMB.seq.\*
  - 18: /cgn2\_6/ptodata/2/pna/US094\_COMB.seq.\*
  - 19: /cgn2\_6/ptodata/2/pna/US095A\_COMB.seq.\*
  - 20: /cgn2\_6/ptodata/2/pna/US095B\_COMB.seq.\*
  - 21: /cgn2\_6/ptodata/2/pna/US095C\_COMB.seq.\*
  - 22: /cgn2\_6/ptodata/2/pna/US096A\_COMB.seq.\*
  - 23: /cgn2\_6/ptodata/2/pna/US096B\_COMB.seq.\*
  - 24: /cgn2\_6/ptodata/2/pna/US097\_COMB.seq.\*
  - 25: /cgn2\_6/ptodata/2/pna/US6000\_COMB.seq.\*
  - 26: /cgn2\_6/ptodata/2/pna/US6001\_COMB.seq.\*
  - 27: /cgn2\_6/ptodata/2/pna/US6002\_COMB.seq.\*
  - 28: /cgn2\_6/ptodata/2/pna/US6003\_COMB.seq.\*
  - 29: /cgn2\_6/ptodata/2/pna/US6004\_COMB.seq.\*
  - 30: /cgn2\_6/ptodata/2/pna/US6005\_COMB.seq.\*
  - 31: /cgn2\_6/ptodata/2/pna/US6006\_COMB.seq.\*
  - 32: /cgn2\_6/ptodata/2/pna/US6007\_COMB.seq.\*
  - 33: /cgn2\_6/ptodata/2/pna/US6008\_COMB.seq.\*
  - 34: /cgn2\_6/ptodata/2/pna/US6009\_COMB.seq.\*
  - 35: /cgn2\_6/ptodata/2/pna/US6010\_COMB.seq.\*
  - 36: /cgn2\_6/ptodata/2/pna/US6011\_COMB.seq.\*
  - 37: /cgn2\_6/ptodata/2/pna/US6012\_COMB.seq.\*
  - 38: /cgn2\_6/ptodata/2/pna/US6013\_COMB.seq.\*
  - 39: /cgn2\_6/ptodata/2/pna/US6014\_COMB.seq.\*
  - 40: /cgn2\_6/ptodata/2/pna/US6015\_COMB.seq.\*
  - 41: /cgn2\_6/ptodata/2/pna/US6016\_COMB.seq.\*
  - 42: /cgn2\_6/ptodata/2/pna/US6017\_COMB.seq.\*
  - 43: /cgn2\_6/ptodata/2/pna/US6018\_COMB.seq.\*

- 44: /cgn2\_6/ptodata/2/pna/US6019\_COMB.seq.\*
- 45: /cgn2\_6/ptodata/2/pna/US6020\_COMB.seq.\*
- 46: /cgn2\_6/ptodata/2/pna/US6021\_COMB.seq.\*
- 47: /cgn2\_6/ptodata/2/pna/US6022\_COMB.seq.\*
- 48: /cgn2\_6/ptodata/2/pna/US6023\_COMB.seq.\*
- 49: /cgn2\_6/ptodata/2/pna/US6024\_COMB.seq.\*
- 50: /cgn2\_6/ptodata/2/pna/US6025\_COMB.seq.\*
- 51: /cgn2\_6/ptodata/1/pna/PCT\_NEW\_COMB.seq.\*
- 52: /cgn2\_6/ptodata/1/pna/US06\_NEW\_COMB.seq.\*
- 53: /cgn2\_6/ptodata/1/pna/US07\_NEW\_COMB.seq.\*
- 54: /cgn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq.\*
- 55: /cgn2\_6/ptodata/1/pna/US09\_NEW\_COMB.seq.\*
- 56: /cgn2\_6/ptodata/1/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	258	100.0	453	19	US-09-528-409-88514	Sequence 88514, A
2	258	100.0	506	24	US-09-726-805-429	Sequence 429, App
3	258	100.0	521	22	US-09-644-869-356	Sequence 356, App
4	258	100.0	521	44	US-60-196-718-2355	Sequence 2355, App
5	258	100.0	529	23	US-09-699-998-4785	Sequence 4785, App
6	258	100.0	617	17	US-09-396-885-430	Sequence 430, App
7	258	100.0	657	19	US-09-522-303-1025	Sequence 1025, App
8	258	100.0	673	23	US-09-698-010-14935	Sequence 14935, A
9	258	100.0	673	23	US-09-699-998-9797	Sequence 9797, App
10	258	100.0	673	24	US-09-710-286-3772	Sequence 3772, App
11	258	100.0	673	24	US-09-716-473-2508	Sequence 2500, App
12	258	100.0	673	24	US-09-721-588-4548	Sequence 4548, App
13	258	100.0	673	55	US-09-726-787-2982	Sequence 2982, App
14	258	100.0	706	23	US-09-699-998-1778	Sequence 1778, App
15	258	100.0	803	23	US-09-699-998-10469	Sequence 10469, A
16	258	100.0	888	23	US-09-652-109-10087	Sequence 9303, App
17	258	100.0	888	23	US-09-652-126-9303	Sequence 9065, App
18	258	100.0	888	23	US-09-652-816-9065	Sequence 2323, App
19	258	100.0	888	24	US-09-716-473-2323	Sequence 1375, App
20	258	100.0	888	24	US-09-716-990-1375	Sequence 4181, App
21	258	100.0	888	24	US-09-721-588-4181	Sequence 2042, App
22	258	100.0	888	24	US-09-726-172-2642	Sequence 1977, App
23	258	100.0	888	24	US-09-726-805-1977	Sequence 2896, App
24	258	100.0	888	55	US-09-726-809-2896	Sequence 9, Appli
25	258	100.0	1228	11	US-08-799-910-9	Sequence 9, Appli
26	258	100.0	1228	12	US-08-825-486-9	Sequence 9, Appli
27	258	100.0	1228	12	US-08-870-434-5	Sequence 5, Appli
28	258	100.0	1228	13	US-08-826-248-9	Sequence 9, Appli
29	258	100.0	1228	13	US-08-925-588-9	Sequence 9, Appli
30	258	100.0	1228	17	US-09-372-044-9	Sequence 46, Appli
31	258	100.0	1308	1	PCT-US00-15136-46	Sequence 2393, App
32	258	100.0	1316	17	US-09-338-425-2393	Sequence 9722, App
33	258	100.0	1316	22	US-09-649-162-9722	Sequence 9454, App
34	258	100.0	1316	23	US-09-652-109-9454	Sequence 10214, A
35	258	100.0	1316	23	US-09-652-128-10214	Sequence 9119, App
36	258	100.0	1316	23	US-09-652-128-10214	Sequence 9799, App
37	258	100.0	1316	23	US-09-699-998-9799	Sequence 2658, App
38	258	100.0	1316	24	US-09-710-280-2658	Sequence 2888, App
39	258	100.0	1316	24	US-09-716-473-2888	Sequence 3900, App
40	258	100.0	1316	24	US-09-721-588-3900	Sequence 2410, App
41	258	100.0	1316	24	US-09-726-171-2410	Sequence 2393, App
42	258	100.0	1316	24	US-09-726-790-2393	Sequence 1703, App
43	258	100.0	1316	24	US-09-726-791-1703	Sequence 1978, App
44	258	100.0	1316	24	US-09-726-805-1978	Sequence 2984, App
45	258	100.0	1316	55	US-09-726-787-2984	Sequence 2984, App

ALIGNMENTS

## RESULT 1

US-09-528-409-88514  
; Sequence 88514, Application US/09528409  
; GENERAL INFORMATION:  
; APPLICANT: Drmanac, Radoje T.  
; APPLICANT: Labat, Ivan  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Dickson, Mark  
; APPLICANT: Jones, Lee W.  
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained  
; FILE REFERENCE: 774  
; CURRENT APPLICATION NUMBER: US/09/528,409  
; CURRENT FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/125,453  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 116231  
; SOFTWARE: Hy-patent.pl Version 3.1  
; SEQ ID NO 88514  
; LENGTH: 453  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)-(453)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-528-409-88514

Query Match 100.0%; Score 258; DB 19; Length 453;  
Best Local Similarity 100.0%; Pred. No. 7.4e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCCGGCCAGCTGCCAGTCCGAGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60  
DB 51 gtccggccagctgccagtcgagaaacccagccaaaggtctctcttctgctg 110  
QY 61 CTCACCATCTGCTTCTGCCAGATCTGATGCTGAAGAGGTGTGCCGGGCCCTGCCT 120  
DB 111 ctccaccatctctctgcagatctgatgctgaagaggtgtgctgctgctgct 170  
QY 121 CCAGAGGAGCCCTTAACGGCGATCTGCGCCAGCTGCTGGACCTCAGACCTTTCTCCAG 180  
DB 171 ccagaggagcccttaacgcgcgcctgctgctgctgctgctgctgctgctgct 230  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACCTTTCTCCAG 240  
DB 231 ccccttaatctgacttcggagccctcgagctgctgctgctgctgctgctgct 290  
QY 241 CAACACCGCGCGCTTC 258  
DB 291 caacacccggcgcccttc 308

## RESULT 2

US-09-726-805-429  
; Sequence 429, Application US/09726805  
; GENERAL INFORMATION:  
; APPLICANT: Gearing, David P.  
; APPLICANT: Holtzman, Douglas A.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; FILE REFERENCE: 1600.2017-001  
; CURRENT APPLICATION NUMBER: US/09/726,805  
; CURRENT FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: 60/168,140  
; PRIOR FILING DATE: 1999-11-30  
; NUMBER OF SEQ ID NOS: 2158  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 429  
; LENGTH: 506  
; TYPE: DNA  
; ORGANISM: Homo sapiens

## US-09-726-805-429

Query Match 100.0%; Score 258; DB 24; Length 506;  
Best Local Similarity 100.0%; Pred. No. 7.5e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCCGGCCAGCTGCCAGTCCGAGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60  
DB 138 gtccggccagctgccagtcgagaaacccagccaaaggtctctcttctgctg 197  
QY 61 CTCACCATCTGCTTCTGCCAGATCTGATGCTGAAGAGGTGTGCCGGGCCCTGCCT 120  
DB 198 ctccaccatctctctgcagatctgatgctgaagaggtgtgctgctgctgctgct 257  
QY 121 CCAGAGGAGCCCTTAACGGCGATCTGCGCCAGCTGCTGGACCTCAGACCTTTCTCCAG 180  
DB 258 ccagaggagcccttaacgcgcgcctgctgctgctgctgctgctgctgctgct 317  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACCTTTCTCCAG 240  
DB 318 ccccttaatctgacttcggagccctcgagctgctgctgctgctgctgctgct 377  
QY 241 CAACACCGCGCGCTTC 258  
DB 378 caacacccggcgcccttc 395

## RESULT 3

US-09-644-869-356  
; Sequence 356, Application US/09644869  
; GENERAL INFORMATION:  
; APPLICANT: Shyjan, Andrew W.  
; APPLICANT: McCarthy, Sean A.  
; APPLICANT: Holtzman, Douglas A.  
; APPLICANT: Monahan, John  
; APPLICANT: Richardson, Jennifer  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; FILE REFERENCE: 1600.1182-001  
; CURRENT APPLICATION NUMBER: US/09/644,869  
; CURRENT FILING DATE: 2000-08-28  
; PRIOR APPLICATION NUMBER: 60/151,062  
; PRIOR FILING DATE: 1999-08-27  
; NUMBER OF SEQ ID NOS: 9708  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 356  
; LENGTH: 521  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-644-869-356

Query Match 100.0%; Score 258; DB 22; Length 521;  
Best Local Similarity 100.0%; Pred. No. 7.5e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCCGGCCAGCTGCCAGTCCGAGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60  
DB 211 gtccggccagctgccagtcgagaaacccagccaaaggtctctcttctgctg 270  
QY 61 CTCACCATCTGCTTCTGCCAGATCTGATGCTGAAGAGGTGTGCCGGGCCCTGCCT 120  
DB 271 ctccaccatctctctgcagatctgatgctgaagaggtgtgctgctgctgctgct 330  
QY 121 CCAGAGGAGCCCTTAACGGCGATCTGCGCCAGCTGCTGGACCTCAGACCTTTCTCCAG 180  
DB 331 ccagaggagcccttaacgcgcgcctgctgctgctgctgctgctgctgctgct 390  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACCTTTCTCCAG 240  
DB 391 ccccttaatctgacttcggagccctcgagctgctgctgctgctgctgctgctgct 450

Best Local Similarity 100.0%; Pred. No. 7.5e-52; Mismatches 0; Indels 0; Gaps 0; Matches 258; Conservative 0;

QY 1 GTCCGGCGCCAGCTGCCAGTTCGAGAACCCAGCCAAAGGCTTCTCTTCTCTG 60  
DB 182 gtccggcgccagctgccagtcgaggaacccagccaaaggctctctctctgctg 241  
QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGCTGTGCCGGCGCCCTG 120  
DB 242 ctccaccatcgctcttctgccagatccctgatggctgaagagggtgtgcccgctccct 301  
QY 121 CCAGAGGACCCCTTAACCGCCATCCTGCGGCCACCCCTGTGTCCCGCTCTCCAG 180  
DB 302 ccagaggagcccttaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 361  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCAGACATTTCTCCAG 240  
DB 362 ccctttaatctgacttcggagccctcgagctacgtctgtggacctcagcacttctccag 421  
QY 241 CAACACCGCGCGCTTC 258  
DB 422 caacaccgcgcccttc 439

RESULT 6

US-09-396-885-430  
; Sequence 430, Application US/09396885  
; GENERAL INFORMATION:  
; APPLICANT: Gearing, David P.  
; APPLICANT: Busfield, Samantha J.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A  
; TITLE OF INVENTION: HUMAN ESOPHAGUS LIBRARY  
; FILE REFERENCE: MLN98-48PM  
; CURRENT APPLICATION NUMBER: US/09/396,885  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: 60/100,459  
; PRIOR FILING DATE: 1998-09-15  
; PRIOR APPLICATION NUMBER: 60/106,444  
; PRIOR FILING DATE: 1998-10-30  
; PRIOR APPLICATION NUMBER: 60/126,842  
; PRIOR FILING DATE: 1999-03-30  
; NUMBER OF SEQ ID NOS: 5360  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 430  
; LENGTH: 617  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(617)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-396-885-430

Query Match 100.0%; Score 258; DB 17; Length 617;  
Best Local Similarity 100.0%; Pred. No. 7.6e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGTTCGCCAGTTCGAGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60  
DB 106 gtccggcgccagctgccagtcgaggaacccagccaaaggctctctctctgctg 165  
QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGCTGTGCCGGCGCCCTG 120  
DB 166 ctccaccatcgctcttctgccagatccctgatggctgaagagggtgtgcccgctccct 225  
QY 121 CCAGAGGACCCCTTAACCGCCATCCTGCGGCCACCCCTGTGTCCCGCTCTCCAG 180  
DB 226 ccagaggagcccttaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 285  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCAGACATTTCTCCAG 240  
DB 286 ccctttaatctgacttcggagccctcgagctacgtctgtggacctcagcacttctccag 345

QY 241 CAACACCGCGCGCTTC 258  
DB 451 caacaccgcgcccttc 468  
RESULT 4  
US-60-196-718-2355/c  
; Sequence 2355, Application US/60196718  
; GENERAL INFORMATION:  
; APPLICANT: Bonazzi, Vivien  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL000456  
; CURRENT APPLICATION NUMBER: US/60/196,718  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7494  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2355  
; LENGTH: 521  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-60-196-718-2355

Query Match 100.0%; Score 258; DB 44; Length 521;  
Best Local Similarity 100.0%; Pred. No. 7.5e-52; Mismatches 0; Indels 0; Gaps 0;  
Matches 258; Conservative 0;

QY 1 GTCCGGCGCCAGTTCGCCAGTTCGAGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60  
DB 400 gtccggcgccagctgccagtcgaggaacccagccaaaggctctctctctgctg 341  
QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGCTGTGCCGGCGCCCTG 120  
DB 340 CTCACCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGCTGTGCCGGCGCCCTG 281  
QY 121 CCAGAGGACCCCTTAACCGCCATCCTGCGGCCACCCCTGTGTCCCGCTCTCCAG 180  
DB 280 CCAGAGGACCCCTTAACCGCCATCCTGCGGCCACCCCTGTGTCCCGCTCTCCAG 221  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCAGACATTTCTCCAG 240  
DB 220 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCAGACATTTCTCCAG 161  
QY 241 CAACACCGCGCGCTTC 258  
DB 160 caacaccgcgcccttc 143

RESULT 5

US-09-699-998-4785  
; Sequence 4785, Application US/09699998  
; GENERAL INFORMATION:  
; APPLICANT: Holtzman, Douglas A.  
; APPLICANT: Gearing, David P.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; TITLE OF INVENTION: THEREFOR  
; FILE REFERENCE: 1600.2008-001  
; CURRENT APPLICATION NUMBER: US/09/699,998  
; CURRENT FILING DATE: 2000-10-30  
; PRIOR APPLICATION NUMBER: 60/162,362  
; PRIOR FILING DATE: 1999-10-29  
; NUMBER OF SEQ ID NOS: 10905  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4785  
; LENGTH: 529  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-699-998-4785

Query Match 100.0%; Score 258; DB 23; Length 529;

Search completed: January 30, 2001, 21:28:30  
Job time: 19992 sec



```

1  TITLE OF INVENTION: Universal Donor Cells
2  NUMBER OF SEQUENCES: 4
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE: Kilpatrick & Cody
5  STREET: 1100 Peachtree Street, Suite 2800
6  CITY: Atlanta
7  STATE: Georgia
8  COUNTRY: U.S
9  ZIP: 30309-4530
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: PatentIn Release #1.0, Version #1.25
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/88/087,007
17 FILING DATE: 19930701
18 CLASSIFICATION: 435
19 ATTORNEY/AGENT INFORMATION:
20 NAME: Babst, Palrea L.
21 REGISTRATION NUMBER: 31,284
22 REFERENCE/DOCKET NUMBER: OMRF135
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: 404-815-6500
25 TELEFAX: 404-815-6555
26 INFORMATION FOR SEQ ID NO: 2:
27 SEQUENCE CHARACTERISTICS:
28 LENGTH: 2847 base pairs
29 TYPE: nucleic acid
30 STRANDEDNESS: single
31 TOPOLOGY: linear
32 MOLECULE TYPE: cDNA
33 HYPOTHEetical: NO
34 ANTI-SENSE: NO
35 ORIGINAL SOURCE:
36 ORGANISM: Homo sapiens
37 IMMEDIATE SOURCE:
38 LIBRARY: GenBank HUMDAF; HUMDAFCL
39 CLONE: Human DAF cDNA
40 FEATURE:
41 NAME/KEY: misc_feature
42 LOCATION: 1..819
43 OTHER INFORMATION: /note= "HUMDAFCL (Promotor
44 OTHER INFORMATION: end of Exon 1, genomic sequ
45 IS-08-087-007-2

```

Query Match 3.3%; Score 40; DB 1; Length 2847;  
Best Local Similarity 51.1%; Pred. No. 0.3;  
Matches 94; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

[illegible]

RESULT 12  
US-08-483-433-2  
; Sequence 2, Application US/08483433  
; Patent No. 6100443  
; GENERAL INFORMATION:

```

1  APPLICANT: Sims et al.
2  TITLE OF INVENTION: Universal Donor Cells
3  NUMBER OF SEQUENCES: 6
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: Patrea L. Pabst
6  STREET: 2800 One Atlantic Center
7  STREET: 1201 West Peachtree Street
8  CITY: Atlanta
9  STATE: Georgia
10 COUNTRY: U.S.
11 ZIP: 30309-3450
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC Compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: PatentIn Release #1.0, Version #1.25
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/08/483,433
19 FILING DATE:
20 CLASSIFICATION: 800
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: 08/087,007
23 FILING DATE: July 1, 1993
24 CLASSIFICATION: 800
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: 07/906,394
27 FILING DATE: June 29, 1992
28 CLASSIFICATION: 800
29 ATTORNEY/AGENT INFORMATION:
30 NAME: Pabst, Patrea L.
31 REGISTRATION NUMBER: 31,284
32 REFERENCE/DOCKET NUMBER: OMRF135c1p2 div
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE: 404-873-8794
35 TELEFAX: 404-873-9794
36 INFORMATION FOR SEQ ID NO: 2:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 2847 base pairs
39 TYPE: nucleic acid
40 STRANDEDNESS: single
41 TOPOLOGY: linear
42 MOLECULE TYPE: cDNA
43 HYPOTHEetical: NO
44 ANTI-SENSE: NO
45 ORIGINAL SOURCE:
46 ORGANISM: Homo sapiens
47 IMMEDIATE SOURCE:
48 LIBRARY: GenBank HUMDAF; HUMDAFC1
49 CLONE: Human DAF cDNA
50 FEATURE:
51 NAME/KEY: misc_feature
52 LOCATION: 1..819
53 OTHER INFORMATION: /note= "HUMDAFC1 (Promotor
54 OTHER INFORMATION: end of Exon 1, genomic sequ
55 US-08-483-433-2

```

```
Query Match      3.3%; Score 40; DB 3; Length 2847;
Best Local Similarity 51.1%; Pred. No. 0.3;
Matches 94: Conservative 0; Mismatches 90; Indels
```

Qy	326	TGCTCCAGAGGAGCGCCCTTAAGCCGGCATCCCTGGCGGCCACCCCTGTGTCTCTCCGTCGTC	385
Db	505	TGAGTCCAGAGGGTGTGCCAGAGCTVCCVCCCTCCCTTCCCTCCCCACTCTCCCCGAGTC	564
Qy	386	TCGAGCGCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTTCAGACATTTCC	445
Db	565	TAGGCCCGGGGTATGACCGCGGAGCCCTCTGACCCGACCTCTGACCACACACAAACCC	624
Qy	446	TCCAGCAACACCCGGCGCCCTTCTACATGTGACTCCCGGCACTCCCCAAAAGAATCCGA	505
Db	625	TACTCCACCCGCTTTGTCTCTCCACCTTGTGTGACGAGAGCCCGACGCCAGACCCCG	684







;; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
;; NUMBER OF SEQUENCES: 52  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Foley & Lardner  
;; STREET: 1800 Diagonal Road, Suite 500  
;; CITY: Alexandria  
;; STATE: VA  
;; COUNTRY: USA  
;; ZIP: 22313-0299  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/232,463  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/935,313  
;; FILING DATE:  
;; APPLICATION NUMBER: EP 91 114 300.6  
;; FILING DATE: 26-AUG-1991  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BENT, Stephen A.  
;; REGISTRATION NUMBER: 29,768  
;; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703)836-9300  
;; TELEFAX: (703)683-4109  
;; TELEX: 899149  
;; INFORMATION FOR SEQ ID NO: 14:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 7218 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; IMMEDIATE SOURCE:  
;; CLONE: pTZ9pt-Fls  
US-08-232-463-14

Query Match 3.6%; Score 43.8; DB 1; Length 7218;  
Best Local Similarity 2.4%; Pred. No. 0.046;  
Matches 9; Conservative 214; Mismatches 156; Indels 0; Gaps 0;

OY 848 TTGGTATTATTACAGCTTTGTGGACTGTGTGAAGCAGACACCTCGGAACCTCCGCGCAAG 907  
Db 1441 TTTGGTACRR 1382  
OY 908 TAGGAGACAAATGGGAGGACTCGGTGGGGAGGAGCTCCCGCTGGGATGAAGTCTG 967  
Db 1381 RRR 1322  
OY 968 GTGGTGGGTCTGAAGTTAGGAGGTGACTGCTCCAGCATCTCAACTCCGCTGTCT 1027  
Db 1321 RRR 1262  
OY 1028 ACTGTGTGAGACTTCGCGGACCATTAGGAATGAGATCCGTGAGATCTTCCATCTCTT 1087  
Db 1261 RRR 1202  
OY 1088 CAAGTCGCTTTAGGTGGCTCCGAGGTAGAGGTTGGGGTTGGGTGGCTGTCCAGGAG 1147  
Db 1201 RRR 1142  
OY 1148 CGACTGTGAGATCGCTAGTATGTTCTGTGAACACAAATAAATTGATTACTGTCAA 1207  
Db 1141 RRR 1082  
OY 1208 AAAAAAATAAATCG 1226  
Db 1081 RRRRRRRRRRRRRATCG 1063

RESULT 6  
US-07-945-283-1/c  
; Sequence 1, Application US/07945283  
; Patent No. 5352596  
; GENERAL INFORMATION:  
; APPLICANT: Cheung, Andrew K.  
; APPLICANT: Wesley, Ronald D.  
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants  
; TITLE OF INVENTION: Involving The EPO and LIT Genes  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis P. Ribando  
; STREET: 1815 No. 5352596th University Street  
; CITY: Peoria  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 61604  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/945,283  
; FILING DATE: 19920911  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ribando, Curtis P.  
; REGISTRATION NUMBER: 27976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 309-685-4011 ext. 513  
; TELEFAX: 309-685-4128  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8438 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: pseudorabies virus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 622..6495  
; FEATURE:  
; NAME/KEY: variation  
; LOCATION: replace(1099, "g")  
; FEATURE:  
; NAME/KEY: variation  
; LOCATION: replace(1267, "t")  
; FEATURE:  
; NAME/KEY: variation  
; LOCATION: replace(1381, "c")  
; FEATURE:  
; NAME/KEY: variation  
; LOCATION: replace(1566, "c")  
; FEATURE:  
; NAME/KEY: variation  
; LOCATION: replace(7010, "g")  
US-07-945-283-1

Query Match 3.4%; Score 42; DB 1; Length 8438;  
Best Local Similarity 47.8%; Pred. No. 0.15;  
Matches 153; Conservative 0; Mismatches 165; Indels 2; Gaps 1;

OY 56 CGCGCCCTCCACCATCCCGGACCCCGGGCTCGGTCTCCATCAGGGCTCCGAGCAGCGTCCGCTCG 114  
Db 6455 CGCGGGGTCCCCCGGTCTCCATCAGGGCTCCGAGCAGCGTCCGCTCGCTCG 6396

Db 421 tacgtcttgagctcagcaacttctctccagcaacccggcgccttcttaactgtgactc 480  
Qy 481 CCCGCACTCCCAAAAGAAATCCGAAACACCAAGAAACACCAAGCGGTACCTGTCGCG 540  
Db 481 cccgcactcccccaaaagagatccgaaaacacacaaagaaacacccagcgctaccctggcg 540  
Qy 541 CGAGAGCGTATCCCACTGGGACTCCCAAGCACTTGAACACTCAGAACACTACAGCGGA 600  
Db 541 cgagagcglatcccccaactgggaactccgagggcaactlgaaactcaagaaactcaagcgga 600  
Qy 601 GAGCGCACCCGCTGCTTGAAGCGGACCGAGCGCACAGACCCAGCGGCGATAGAGACC 660  
Db 601 gagcgcaacccggctgtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 660  
Qy 661 GAGGCGACAGCCAGCTGGGCTAGGCCCGGTGGGAGGAGAGCGTCTTAATTTATTTCT 720  
Db 661 gagggcaagccagctggggctagggcgagcgagcgagcgagcgagcgagcgagcgagcg 720  
Qy 721 TATTGCTCTTAATTAATTTATATATATATATATATATATATATATATATATATAT 780  
Db 721 tatgctcccaataataattatattatattatattatattatattatattatattatatt 780  
Qy 781 GTACCTATATATTTTAACTATGCAAGGCTGTGAGATGTTCCCTGCTGCTAAATGC 840  
Db 781 gtacgtaatattatttaactatgcgaagggtgagatgttccctctctgtgtaaatgc 840  
Qy 841 AGGTCTCTTGTGTTATTTAGCTTTGTGGACTGTGTGGAGGAGGAGGAGGAGGAGGAG 900  
Db 841 aggtctcttgglatlattagactttgtggagactgtgtggagagagagagagagagag 900  
Qy 901 GCGAAGTAGGAGAGAAATCGGAGGACTCGGGTGGGGAGGAGGAGGAGGAGGAGGAGGAG 960  
Db 901 ggcgaagtaggagagagaaatgggagagactcgggtgggagagagagagagagagagag 960  
Qy 961 AGCTCTGGGCTGGGCTTAAGCTTTAGGAGGCTGACTGCTCCTCAGCATCTCAACTCG 1020  
Db 961 agctctgggtgggtgggtgggtgggtgggtgggtgggtgggtgggtgggtgggtgggtgg 1020  
Qy 1021 TGTGCTACTGTGTGAGACTTCGGGGGACCATTAGGAATGAGATCCGTGAGATCCTTCCA 1080  
Db 1021 tctgctactgtgtgagacttcggcgagaccattaggaalagagatccglagatccctcca 1080  
Qy 1081 TCTTCTGAAGTCCGCTTTAGGCTGCTGCGAGGTAGAGGTGGGGTGGTGGGCTGT 1140  
Db 1081 tctcttgaagtcgccttttaggttgctgagaggtagaggggtgggggtgggtgggtgggt 1140  
Qy 1141 CACGAGGAGCTGTGAGATCCGCTAGTATGTTCTGTGAACACAAATAAATTTGATTTAC 1200  
Db 1141 cacygagcagctgtcagatccctagatcgctatgtctctgtgaacacaaataaaattgattac 1200  
Qy 1201 TGTCAAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 1228  
Db 1201 tgtcaaaaaaaaaaaaaaaaaaactcgag 1228

RESULT 4

US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE: US/07/935,313  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpl-F1s  
; US-08-232-463-14

Query Match 4.1%; Score 49.8; DB 1; Length 7218;  
Best Local Similarity 1.6%; Pred. No. 0.0012;  
Matches 6; Conservative 220; Mismatches 147; Indels 0; Gaps 0;  
Qy 97 CCTGAGATCTTACCTTCGACCCCTCTCCGAGAGCCGCGGCGCCCTGCGGCGGCGCC 156  
Db 1059 CTTGCGATTT 1118  
Qy 157 AGGCCTCTCGCGGACCGAAGCGAGCGAGGCTTCTTACCTCGAGTGTCTCGG 216  
Db 1119 YY 1178  
Qy 217 CGCCAGCTGCCAGTCGAGAACCCAGCAACCCAGCAACCCAGCAACCCAGCAACCC 276  
Db 1179 YY 1238  
Qy 277 ATGCTCTTCTCCAGATCCTGATGCTGAAGAGGTGTGCGGCGGCGCCCTGCCCTCC 336  
Db 1239 YY 1298  
Qy 337 GAGCGCCCTAACGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTGCTCGAGCCCTTT 396  
Db 1299 YY 1358  
Qy 397 AATCTGACTCGGAGCCCTCGGACTACGCTCTGACCTCTGACACTTTCTCTCCAGCAAC 456  
Db 1359 YY 1418  
Qy 457 CCGCGCGGCTTCT 469  
Db 1419 YYYYYYYYYYYYYY 1431

RESULT 5

US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.



US-08-826-246-9

Query Match 100.0%; Score 1228; DB 3; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTCACCTCCAGCTCCAGCCGACCATGACCATCTGCGAGCCCGCCGAGCCCGGCC 60  
Db 1 ATGTCACCTCCAGCTCCAGCTCCAGCCGACCATGACCATCTGCGAGCCCGCCGAGCCCGGCC 60

Qy 61 CCCCTCCACCATCCCGGACCCCGCGGGCTCCGGCTCTGAGATCTCTTCACTCTCGACCT 120  
Db 61 CCCCTCCACCATCCCGGACCCCGCGGGCTCCGGCTCTGAGATCTCTTCACTCTCGACCT 120

Qy 121 CTCCTCCGAGCTCCAGCGGGCCCTGCGCGGGCCCGCCAGCCCTCTGCGGGCCAGCAAG 180  
Db 121 CTCCTCCGAGCTCCAGCGGGCCCTGCGCGGGCCCGCCAGCCCTCTGCGGGCCAGCAAG 180

Qy 181 CGCAGCCGAGGTTCTTCTACCTTCGAGTGGTCCGGGCCAGCTGCGAGTCCGAGAACCG 240  
Db 181 CGCAGCCGAGGTTCTTCTACCTTCGAGTGGTCCGGGCCAGCTGCGAGTCCGAGAACCG 240

Qy 241 AACCCAGCCAAAAGGCTTCTTCTGCTGCTCACCATCTCTTCTGCCAGATCTCTGATG 300  
Db 241 AACCCAGCCAAAAGGCTTCTTCTGCTGCTCACCATCTCTTCTGCCAGATCTCTGATG 300

Qy 301 GCTGAAGAGGTTGCGCGGGCCCTGCTCCAGAGGACCCCTTAACGCCGATCCCTG 360  
Db 301 GCTGAAGAGGTTGCGCGGGCCCTGCTCCAGAGGACCCCTTAACGCCGATCCCTG 360

Qy 361 CGCGCCACCTCTGTCCTCCCGCTCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGAC 420  
Db 361 CGCGCCACCTCTGTCCTCCCGCTCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGAC 420

Qy 421 TACCTCTGAGCTTCAGCTTCTCTCCAGCAACACCCCGCCCTTCTTAACCTGACTC 480  
Db 421 TACCTCTGAGCTTCAGCTTCTCTCCAGCAACACCCCGCCCTTCTTAACCTGACTC 480

Qy 481 CCGCCTCTCCAAAAGAAATCGAAACCAACCAAGAACACAGCGCTTACCTGGTGG 540  
Db 481 CCGCCTCTCCAAAAGAAATCGAAACCAACCAAGAACACAGCGCTTACCTGGTGG 540

Qy 541 CGAGAGCTATCCCACTCGGACTTCCGAGGCACTTGAACCTCAGAACACTACAGCGGA 600  
Db 541 CGAGAGCTATCCCACTCGGACTTCCGAGGCACTTGAACCTCAGAACACTACAGCGGA 600

Qy 601 GACCCACCCGGTCTTGAAGCGGACCGAGCGCACAGACCGAGCGCATACAGAGACC 660  
Db 601 GACCCACCCGGTCTTGAAGCGGACCGAGCGCACAGACCGAGCGCATACAGAGACC 660

Qy 661 GAGCAGACCCAGCTGGGGCTAGGCGCGTGGAGAGGAGCGTCTTAATTTATTTCT 720  
Db 661 GAGCAGACCCAGCTGGGGCTAGGCGCGTGGAGAGGAGCGTCTTAATTTATTTCT 720

Qy 721 TATTGCTCTTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTCT 780  
Db 721 TATTGCTCTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTCT 780

Qy 781 GTACCTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTCT 840  
Db 781 GTACCTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTCT 840

Qy 841 AGGTCTCTGTTATTTATTTAGCTTTGTTGGACTGTTGGAGTGGTGGAGGAGGAGTGG 900  
Db 841 AGGTCTCTGTTATTTATTTAGCTTTGTTGGACTGTTGGAGTGGTGGAGGAGGAGTGG 900

Qy 901 GCGAAAGTAGGAGAAATGCGGAGGACTTCGGGTGGGAGGAGGAGGAGGAGGAGGAGG 960  
Db 901 GCGAAAGTAGGAGAAATGCGGAGGACTTCGGGTGGGAGGAGGAGGAGGAGGAGGAGG 960

Qy 961 AAGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1020  
Db 961 AAGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1020

Db 961 AAGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1020  
Qy 1021 TCTGTCTACTGTGTGAGACTTCCGGCGGACCATTTAGGAATGAGATCCGTGAGATCTTCCA 1080  
Db 1021 TCTGTCTACTGTGTGAGACTTCCGGCGGACCATTTAGGAATGAGATCCGTGAGATCTTCCA 1080  
Qy 1081 TCTTCTTGAAGTCGCCCTTTAGGCTGCTCGAGGTAGAGGTTGGGGGTTGGTGGGCTGT 1140  
Db 1081 TCTTCTTGAAGTCGCCCTTTAGGCTGCTCGAGGTAGAGGTTGGGGGTTGGTGGGCTGT 1140  
Qy 1141 CACGGAGCAGCTGTGAGATCCCTTAGTATGTTCTGTGAACACAAATAAATTTGATTTAC 1200  
Db 1141 CACGGAGCAGCTGTGAGATCCCTTAGTATGTTCTGTGAACACAAATAAATTTGATTTAC 1200  
Qy 1201 TGTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1228  
Db 1201 TGTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1228

RESULT 2  
US-08-944-495-9  
; Sequence 9, Application US/08944495  
; Patent No. 6087477  
; GENERAL INFORMATION:  
; APPLICANT: Faib, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 2.0  
; CURRENT APPLICATION DATA: US/08/944.495  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/799,910  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-067-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)7909090  
; TELEFAX: (212)8699741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1228 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 1...468  
; OTHER INFORMATION:  
; US-08-944-495-9

Query Match 100.0%; Score 1228; DB 3; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 15:55:18 ; Search time 49.1 seconds  
(without alignments)  
4030.649 Million cell updates/sec

Title: US-08-799-910-9

Perfect score: 1228

Sequence: 1 ATGTGTCACCTCGCAGCTG.....AAAAAATTCGAG 1228

Scoring table: IDENTITY\_NUC \*

Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1228	100.0	1228	3	US-08-826-246-9
2	1228	100.0	1228	3	US-08-944-495-9
3	1228	100.0	1228	3	US-09-126-640-5
4	49.8	4.1	7218	1	US-08-232-463-14
5	43.8	3.6	7218	1	US-08-232-463-14
6	42	3.4	8438	1	US-07-945-283-1
7	41.2	3.4	530	3	US-08-758-662-4
8	41.2	3.4	4524	2	US-08-845-998-7
9	41.2	3.4	4524	3	US-09-206-537-7
10	40.6	3.3	152331	3	US-09-128-155-16
11	40	3.3	2847	3	US-08-087-007-2
12	40	3.3	2847	3	US-08-483-433-2
13	40	3.3	2847	4	PCT-US92-05920-2
14	39.8	3.2	1120	3	US-09-030-613-1
15	38.6	3.1	6354	3	US-09-058-389A-5
16	38.4	3.1	6803	3	US-08-665-259-19
17	38.4	3.1	6803	3	US-08-762-500-19
18	37.6	3.1	170	3	US-09-058-389A-13
19	37.6	3.1	4425	2	US-08-749-169A-1
20	37.6	3.1	4425	2	US-09-130-032A-1
21	37.4	3.0	1420	2	US-08-909-965C-3
22	37.2	3.0	371	1	US-08-664-596B-25
23	37.2	3.0	371	2	US-08-739-775-3
24	37.2	3.0	20235	1	US-07-642-734C-3
25	37.2	3.0	20235	3	US-08-439-009A-3
26	36.6	3.0	1931	2	US-09-130-114-2
27	36.4	3.0	702	1	US-08-458-568A-3
28	36.4	3.0	1841	2	US-08-820-521-1

Sequence 1, Appli  
Sequence 5, Appli  
Sequence 11, Appli  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 213, App  
Sequence 11, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli

#### ALIGNMENTS

RESULT 1  
US-08-826-246-9  
; Sequence 9, Application US/08826246  
; Patent No. 6048709  
; GENERAL INFORMATION:  
; APPLICANT: Faib, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/826,246  
; FILING DATE: 28-MAR-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/799,910  
; FILING DATE: 13-FEB-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/011,787  
; FILING DATE: 16-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-078-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)7909090  
; TELEFAX: (212)8699741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1228 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 1..468  
; OTHER INFORMATION:

GenCore version 4.5  
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 16:08:34 ; Search time 77.57 Seconds  
(without alignments)  
5947.064 Million cell updates/sec

Title: US-08-799-910-9  
Perfect score: 1228  
Sequence: 1 ATGCTCACTCGCAGCTG.....AAAAAATCGAG 1228

Scoring table: IDENTITY\_NUC %  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*\*  
1: /cgn2\_2/gcgdata/geneseq/geneseq/NA1980.DAT:\*  
2: /cgn2\_2/gcgdata/geneseq/geneseq/NA1981.DAT:\*  
3: /cgn2\_2/gcgdata/geneseq/geneseq/NA1982.DAT:\*  
4: /cgn2\_2/gcgdata/geneseq/geneseq/NA1983.DAT:\*  
5: /cgn2\_2/gcgdata/geneseq/geneseq/NA1984.DAT:\*  
6: /cgn2\_2/gcgdata/geneseq/geneseq/NA1985.DAT:\*  
7: /cgn2\_2/gcgdata/geneseq/geneseq/NA1986.DAT:\*  
8: /cgn2\_2/gcgdata/geneseq/geneseq/NA1987.DAT:\*  
9: /cgn2\_2/gcgdata/geneseq/geneseq/NA1988.DAT:\*  
10: /cgn2\_2/gcgdata/geneseq/geneseq/NA1989.DAT:\*  
11: /cgn2\_2/gcgdata/geneseq/geneseq/NA1990.DAT:\*  
12: /cgn2\_2/gcgdata/geneseq/geneseq/NA1991.DAT:\*  
13: /cgn2\_2/gcgdata/geneseq/geneseq/NA1992.DAT:\*  
14: /cgn2\_2/gcgdata/geneseq/geneseq/NA1993.DAT:\*  
15: /cgn2\_2/gcgdata/geneseq/geneseq/NA1994.DAT:\*  
16: /cgn2\_2/gcgdata/geneseq/geneseq/NA1995.DAT:\*  
17: /cgn2\_2/gcgdata/geneseq/geneseq/NA1996.DAT:\*  
18: /cgn2\_2/gcgdata/geneseq/geneseq/NA1997.DAT:\*  
19: /cgn2\_2/gcgdata/geneseq/geneseq/NA1998.DAT:\*  
20: /cgn2\_2/gcgdata/geneseq/geneseq/NA1999.DAT:\*  
21: /cgn2\_2/gcgdata/geneseq/geneseq/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1228	100.0	1228	21	250711	Nucleotide sequenc
2	1226.4	99.9	1228	18	T94471	Human Fchd605 gene
3	420.8	34.3	427	21	280483	Human colon cancer
4	371.6	30.3	411	20	V86227	EST clone OA17. H
5	167.2	13.6	193	20	V89271	EST clone BY66. H
6	151	12.3	297	15	Q77534	Human genome fragm
7	70.4	5.7	267	15	Q77554	Human genome fragm
8	57.4	4.7	1000	21	A02484	Human colon cancer
9	55.2	4.5	1218	21	A02488	Human colon cancer
10	52.2	4.3	1337	20	217263	Human gene express
11	49.6	4.0	1126	21	A02538	Human colon cancer
12	48.2	3.9	53	16	T20789	Human gene signatu

13	47.8	3.9	4228	21	251683	Human cyclic nucle
14	47.2	3.8	114955	20	X53491	Human adenosine A1
15	44.8	3.6	1593	21	A02504	Human colon cancer
16	44.2	3.6	2375	21	A34740	Human adenosine re
17	44.2	3.6	4475	21	A34741	Human adenosine re
18	42.8	3.5	1127	21	A02477	Human colon cancer
19	42.4	3.5	4356	14	Q37543	Cardiac adenylyl c
20	42.4	3.5	4356	16	Q95540	Cardiac adenylyl c
21	42	3.4	8438	15	Q73500	DNA encoding Pseud
22	42	3.4	34094	20	230163	Complete nucleotid
23	41.8	3.4	1459	21	A02528	Human colon cancer
24	41.6	3.4	989	21	A02539	Human colon cancer
25	41.4	3.4	1275	11	Q01620	Chicken beta-actin
26	41.4	3.4	1344	12	Q13318	Expression vector
27	41.4	3.4	1345	11	Q03062	Hybrid promoter of
28	41.2	3.4	1102	20	216923	Human gene express
29	41.2	3.4	4524	20	V33912	Nucleotide sequenc
30	41	3.3	114955	20	X53491	Human adenosine A1
31	40.8	3.3	1966	20	X61220	Mouse DNA demethyl
32	40.6	3.3	1017	20	217058	Human gene express
33	40.4	3.3	300	20	214922	Human gene express
34	40.4	3.3	2179	15	Q73117	Alpha 2, 3-sialyl
35	40.4	3.3	2232	15	Q77739	Human alpha-2,3-si
36	40	3.3	1523	20	217507	Human gene express
37	40	3.3	1523	20	217479	Human gene express
38	39.8	3.2	1120	19	V60339	cDNA sequence of f
39	39.8	3.2	2188	20	277506	Human ovarian tumo
40	39.8	3.2	3198	20	X02974	Human IL-1ra BAC c
41	39.6	3.2	954	20	X30342	DNA encoding a hum
42	39.6	3.2	2681	20	X58429	Thermophilus therm
43	39.2	3.2	1804	20	X61218	Human DNA demethyl
44	38.8	3.2	749	20	216153	Human gene express
45	38.8	3.2	1312	20	217264	Human gene express

ALIGNMENTS

RESULT 1	
ID 250711	standard; DNA; 1228 BP.
AC 250711;	
DT 31-MAY-2000	(first entry)
XX	Nucleotide sequence of human fchd605 gene.
DE	fchd605 gene; human; cardiovascular disease; oncogenic disorder;
KW	diabetic retinopathy; fibroproliferative disorder; atherosclerosis;
KW	TGF-beta signalling pathway; TGF; Transforming growth factor;
KW	pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;
KW	vascularisation; cytostatic; antidiabetic; ophthalmological; ds.
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	1..471
FT	/*tag= a
FT	/product= "fchd605 protein"
XX	
PN	WO200006206-A1.
XX	
PD	10-FEB-2000.
XX	
PF	30-JUL-1999; 99WO-US17394.
XX	
PR	30-JUL-1998; 98US-0126640.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Falb DA;
XX	

DR WPI: 2000-205414/18.  
DR P-PSDB: Y45017.

XX Identifying substances for ameliorating symptoms of fibroproliferative  
PT diseases or oncogenic related disorders  
XX Examples: Fig 5; 214pp; English.

CC The patent discloses methods for the treatment and diagnosis of  
CC cardiovascular diseases by novel human genes which are differentially  
CC expressed in different cardiovascular disease states. Compositions which  
CC can modify TCP-beta signalling pathway are identified by screening.  
CC These are used therapeutically to treat fibroproliferative and oncogenic  
CC disorders, especially TGF (transforming growth factor)-beta related  
CC disorders, including diabetic retinopathy, arteriosclerosis, pancreatic  
CC cancer, angiogenesis, inflammation, fibrosis, tumour growth and  
CC vascularisation. The present sequence is fchd605 gene which is up  
CC -regulated in monocytes treated with oxidised LDL (low density  
CC lipoprotein) can be used to design cardiovascular disease treatment  
CC strategies. Depending on whether the up-regulation has a pathogenic or  
CC protective effect treatment methods can be designed to increase or  
CC decrease the activity of the protein product of the gene.

SO Sequence 1228 BP; 265 A; 361 C; 341 G; 261 T; 0 other;

Query Match 100.0%; Score 1228; DB 21; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATCTGTCACATCTCCAGCTGCCACCGACCATGACCATCTCTGAGGCGCCCGCCGCGCC 60  
DB 1 algtgtcaetctcgcagctgcccacccgacatgacatcccgagggcccccgccgccc 60  
OY 61 CCTTCACCATCCGGGACCCCGGGGCTCCGCTCGTGTGAGATCTTCACCTTCGACCTT 120  
DB 61 cctctcaaceaaccgugaccccggggctccggtcctgagatctctccttcgacct 120  
OY 121 CTCGGGAGCCCGACGGCCCTGCGGGCGCCCGCCAGCGCTCTGAGGCGCCCGCCGCGCC 180  
DB 121 ctccggagcccgagcgccctcgccggcgcccgcccgcccgcccgcccgcccgccg 180  
OY 181 CGAGCGCGAGGCTTCTACCTCGAGTGGTCCGCGCGCGAGTGGCGAGTCCGAGGACCG 240  
DB 181 cgcagcgcaggggtctctacacctcgagtgglccggcgccagctgcagtcgaggaaccg 240  
OY 241 AACCCAGCCAAAAGGCTTCTCTTCTGCTGCTCACCATCGTCTTCTGCGAGATCTCTGATG 300  
DB 241 aacccagccaaaaggcttctcttctgtctgctcaccatcgctctctgcccagatcctgtg 300  
OY 301 GCTGAAGAGGGTGTGCGGCGCCCTGCTCCAGAGGAGCGCCCTAACCGCCGATCCCTG 360  
DB 301 gctgaagaggggtgtgcccgcgcccctgctccagagagcgcccctaaagccgcatccctg 360  
OY 361 CGCCCGACCCCTGTGTCGCCGCTCTCGAGCCCTTTAAATCTGACTTCGAGGCGCTCGGAC 420  
DB 361 ggcggccacccctgtgtcccccgtccctcgagccctttaaactgacttcggagccctcgac 420  
OY 421 TAGCCTCTGACCTCAGCACTTTCTCTCCAGAACACCGCCGCGCTTCTTAACCTGTGACTC 480  
DB 421 tagcctctgagacctcagcacttctctccgagcaacccgcccgtcttctaaactgtgactc 480  
OY 481 CCGCGACTCCCAAAAGAAATCCGAAAACACAAAGAACACACGAGCGGTACCTGTGTCGG 540  
DB 481 cccgcaccccccaaaagaatccgaaaaaaccaaaagaacacagggcgatcctgtgctg 540  
OY 541 CGACAGCCGTATCCCCAATCGGACTTCCGAGGCACTTGAACCTCAGAACACTACACCGGA 600  
DB 541 cgagagcgtatcccccaactgggacttccgagggcaacttgaactcagaaactacacagcga 600  
OY 601 GACGCCACCCGGTGTGTTAGGGGGGACCGGAGCGCACAGACCGGCGCATAGAGACC 660  
DB 601 gacgccacccgggtgtgtttaggggggacccgagcgacagagacccgagcgcatagagacc 660

OY 661 GAGGCACAGCCCGAGCTGGGCTAGGCCGCTGGGAAGAGAGCGTCTTAATTTATTTCT 720  
DB 661 gaggcacagcccgagctgggctagggcggcggagggagagcgcglaaltlaltlct 720  
OY 721 TATTGCTCTTAATAATATATTTATATTTATTTATTTATTTATTTATTTATTTATTT 780  
DB 721 tattgctcttaataatattattattattattattattattattattattattattatt 780  
OY 781 GTACGTAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 840  
DB 781 gtacgtaatattattattattattattattattattattattattattattattattatt 840  
OY 841 AGGTCCTCTTGGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 900  
DB 841 aggtctcttggatattattattattattattattattattattattattattattattatt 900  
OY 901 GCCAAGTAGGAGAGAAATGGGAGGACTCGGGTGGGGAGGAGCGTCCCGGCTGGGATG 960  
DB 901 gccaaagtaggagagaaatgggagggactcgggtggggagagcgtcccggtggatg 960  
OY 961 AAGTCTGTGGTGGTTCGTTAAGTTTAGGAGGTGACTGCTCTCCAGCATCTCAACTCCG 1020  
DB 961 aagtctgtgtgggttcgtaagtttagggaggtgactcgtctccagcatctcaactccg 1020  
OY 1021 TCTGTCTACTGTGTGAGACTTCGGCGACCATTTAGGAATGAGATCCGTGAGATCTCTTCCA 1080  
DB 1021 tctgtctactgtgtgagacttcggcgacctcggcgacctaggaaatgagatccgtgagatctctcca 1080  
OY 1081 TCTTCTTGAAGTCGGCTTTAGGGTGGCTCCGAGGTAGAGGTTGGGGCTTGGGCTGT 1140  
DB 1081 tcttcttgaagtcggcttttaggggtggctcgcaggttagaggggttgggtgggtgtgt 1140  
OY 1141 CACGAGGCGACTGTGAGATCGCTAGTATGTTCTGTGAACAATAAATTTGATTATAC 1200  
DB 1141 cagcgagcgactgtcgagatcgctagatgttctgtgaacaaataaattgattattac 1200  
OY 1201 TGCTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1228  
DB 1201 tgcataaaaaataaataaataaataaataaataaataaataaataaataaataaataa 1228

RESULT 2  
T94471  
ID T94471 standard; cDNA; 1228 BP.

XX T94471;  
XX T94471;  
DT 03-MAR-1998 (first entry)

XX Human Fchd605 gene differentially regulated in monocytes.  
DE Fchd602 gene; differential expression; monocyte; human;  
KW foam cell; cardiovascular disease; atherosclerosis; ischemia;  
KW reperfusion; hypertension; restenosis; arterial inflammation;  
KW therapy; diagnosis; drug screening; marker; ss.

OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 1..471  
FT /tag= a

XX WO9730065-A1.  
XX 21-AUG-1997.  
XX 14-FEB-1997; 97WO-US02291.  
XX 13-FEB-1997; 97US-0799910.  
XX 16-FEB-1996; 96US-0011787.  
PA (MIL-) MILLENNIUM PHARM INC.

XX	
PI	Falb DA;
XX	
XX	WPI; 1997-424966/39.
DR	P-PSOR; W36006.
DR	
XX	
PT	New genes differentially expressed in cardiovascular disease - used
PT	for diagnosis, drug screening and treatment of cardiovascular
PT	disease, e.g. atherosclerosis, restenosis, hypertension, etc
XX	
PS	Claim 1: Fig 5; 163pp; English.
PS	
XX	
CC	Fchd605 is a novel human gene that is up-regulated in monocytes
CC	treated with oxidised low density lipoproteins (LDL). Differential
CC	display was used to detect genes that are differentially expressed
CC	in monocytes treated so as to simulate the conditions under which
CC	foam cells develop during atherogenesis. Both fchd605 and fchd602
CC	(see T94470) are up-regulated under the disease condition of
CC	treatment with oxidised LDL. The fchd605 gene product (see W36006)
CC	has sequence similarity to mouse gly96. The discovery of the
CC	up-regulation of these 2 genes provides a fingerprint profile,
CC	e.g. markers for the study of cardiovascular diseases, including
CC	atherosclerosis, ischaemia/reperfusion, hypertension, restenosis
CC	and arterial inflammation. Methods are provided for the diagnosis,
CC	monitoring in clinical trials, screening for therapeutically
CC	effective compounds, and treatment of cardiovascular diseases based
CC	on discoveries regarding the expression patterns of novel genes
CC	fchd531 (see T94467), fchd540 (see T94468), fchd545 (see T94469),
CC	fchd602 and fchd605.
XX	
XX	
XX	Sequence 1228 BP; 265 A; 362 C; 340 G; 261 T; 0 other:
XX	

Query Match	99.9%	Score 1226.4	DB 18	Length 1228	
Best Local Similarity	99.9%	Pred. No. 0			
Matches 1227	Conservative	0	Mismatches 1	Indels 0	Gaps 0
QY	1	ATGTGTCACCTCGCAGCTGCCACCGACCATGACCATCTCTGCAGGCCCCGACCCCGGCC	60		
Db	1	atgtgtcaactctcgagctgccaacccagcatgacctctctgagggccccgcagcccgcc	60		
QY	61	CCCTCCACCATCCCGGGACCCGGCGGGTCTCGGTCTTGAGATCTTCACTTCGACCT	120		
Db	61	ccctcaccatccgggacccccggggtctcggtcttgagatcttcaacctctgacct	120		
QY	121	CTCCGGACCCCGCAGCGGCCCTTCGCGGGGCCCCAGCGCCTCTCCGGGCACCGAAAG	180		
Db	121	ctccggagccnagcgggccccctgcgggccccagcgctctcgggcaacgaaag	180		
QY	181	CGCAGCGCAGGGTTCTCTACCCTCGAGTCGTCGGCGGCCAGTCGCAGTCGAGGAACGG	240		
Db	181	cgcagccyaggggttctcaacctcgagtggtccggcgccagctgcagtcgagaaatcg	240		
QY	241	AACCCAGCCAAAGGCTTCTTTTCTGCTGCTCACCATCGTCTTCTGCCAGATCTCTGATG	300		
Db	241	aaccagccaaaggcttcttcttctgtctgcaccatcgcttcttcgagatctctgatg	300		
QY	301	GCTGAAGAGCTGTGTCGGCGGCCCTGCTTCGAGAGAGCGCCCTAACGCCGCATCCCTG	360		
Db	301	gctgaagagggtgtgtccgggcccctgcttccagagagcgccctlaacgcgcataccctg	360		
QY	361	GGCCACACCGCTGTGTCGCCCGTCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGAC	420		
Db	361	ggccccacccctgtgtcccccgctctcgagccctttaaactgacttctggagccccctcgac	420		
QY	421	TACGCTCTGGACCTTAGACTTTTCTCCAGAACACCCCGGCCCTTCTTAAGTGTGACTC	480		
Db	421	tacgctctlgacctcagaccttctctccagacaacccccggcgcccttctaactgtlgactc	480		
QY	481	CCCGACCTCCCAAAAGATCCGAAACACCAAGAACACACAGAGCGGTACCTGGTGGC	540		
Db	481	cccgacctccccaaagatccgaaaacacccaaagaaacaccagccagggcgctacctggtcg	540		

QY	541	CGAGAGCGTATCCCCAACTGGGACTTCGAGGCAACTTGAACACACATACACGCGA	600
Db	541	cgagagcgtatcccccaactcggaacttcgagggcaacttgaactcagaaacactacagcga	600
QY	601	GACGCCACCCGGTGTGTAGCGGGGACCGAGGGCGCACAGAGACCGGAGCGCATACAGACC	660
Db	601	gagccaccccggtgcttgagggcggaacggagggcgacagagaccgagggcgca.lagagacc	660
QY	661	GAGGCACAGCCCACTGGGGCTAGGCCCGGTGGGAAGAGAGAGCGTCTGTTAAATTATTTCT	720
Db	661	gaggcacagccagctggtgggctagggcccggtgggaagagagagcgtcgtaattatttct	720
QY	721	TATTCTCTCTTAATTAATATTATATGTATTATGTACGTCTCTTCTAGGTGATGAGATGT	780
Db	721	tattgctcctaataacattattatgtatttatgtacgtctcttctcctcctaggtgtagggagatg	780
QY	781	GTACGTAATATTTATTTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTGCTCAAAATGC	840
Db	781	glacgtaataattattttaacttatgcaaggglgtgagatgtlccctcctgctglglaaalg	840
QY	841	AGGTCTCTTGTTATTTATTTAGCTTTGTGGGACTGTGTGAAGCAGACACCTTGAACATGC	900
Db	841	aggtcctcttggtatttatttgacttttgggactggtggaagcaggacacacctggaactgc	900
QY	901	GGCAAAAGTAGGAGAAGAAATGGGAGGAGCTCGGGTGGGGAGGAGCGTCCCGCTGGGATG	960
Db	901	ggcaaaagtaggagagaagaaa.tggggagagacctcgggtggggagagcgtcccgctgggag	960
QY	961	AAGTCTGGTGTGGGTGCTTAAGTTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCCG	1020
Db	961	aagtctgggtgggtcgtaagttaggaggtagctgcatctccacagcatctcaacilccg	1020
QY	1021	TCTGCTACTGTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCTTCCA	1080
Db	1021	tctgctactgtgtgagacttcggcggaacca.taggaaatgagatcgtgagatcctctcca	1080
QY	1081	TCTTCTTGAAGTCGCCTTTTAGGGTGGCTCGGAGGTAGAGGGTTGGGGGTTGCTGGGCTGT	1140
Db	1081	tcttcttgaagtcgccttttagggtggctcgcaggtagagggttgggggttgggtggcgtgt	1140
QY	1141	CACGAGCGGACTGTCGAGATCGGCTAGTATGTTCTGTGAACACAAATAAAATTGATTTAC	1200
Db	1141	cacggagcgactcagatcgactagtcgtctgtgacacacataaaaaatgattttac	1200
QY	1201	TGTCAAAAAATAAAAAAATAAACTCGAG	1228
Db	1201	tgtcaaaaaataaaaaaataaaactcgag	1228
RESULT	3		
280483/c			
ID	280483	standard; cDNA: 427 BP.	
AC	280483;		
XX			
DT	07-APR-2000	(first entry)	
DE		Human colon cancer cell line SW480 cDNA clone SEQ ID NO:567.	
KW		Human; gene expression product; diagnosis; tumour; colon cancer;	
KW		colorectal adenocarcinoma; cell line SW480; cell proliferation;	
KW		cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;	
KW		hyperplasia; ds.	
OS		Homo sapiens.	
XX			
PN		WO9964576-A2.	
PD		16-DEC-1999.	
PF		09-JUN-1999; 99WO-IB01062.	
XX		10-JUN-1998; 98US-0088801.	
PR			





QY 824 CCTCTGCTGTAATCGAGGTCTCTTGGTATTATTGAGCTTGTGGGACTGGTGAAGC 883  
DB 315 ccncctgctgtaaatcgaggtctcttgggtattattgagcttgtgggactggtggaagc 374  
QY 884 AGCACACCTTGGAACTCGCGCAAGTAGGAGA 914  
DB 375 aggacacctggaactgcccgaagtaggcga 405

RESULT 5  
V89271/C  
ID V89271 standard; cDNA; 193 BP.  
XX  
AC V89271:  
DT  
DE 15-FEB-1999 (first entry)  
DE EST clone BY66.  
XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;  
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
KW gene therapy; ss.  
XX Homo sapiens.  
XX  
XX WO9845436-A2.  
XX  
PD 15-OCT-1998.  
XX  
PF 10-APR-1998; 98WO-US06955.  
XX  
PR 10-APR-1997; 97US-0838821.  
XX  
PA (GEM ) GENETICS INST INC.  
XX  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
PI Racie LA, Spaulding V, Treacy M;  
XX WPI; 1999-070077/06.  
XX New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries.  
XX  
PS Claim 1: Page 164; 618pp; English.  
XX The present sequence represents a human expressed sequence tag (EST).  
CC The polynucleotide, which is a secreted EST, and the encoded protein  
CC are predicted to have useful biological activities which would make  
CC them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals, although no supporting data is  
CC given. Suggested activities include nutritional activity, immune  
CC stimulating or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
XX  
SQ Sequence 193 BP; 35 A; 39 C; 68 G; 51 T; 0 other;

Query Match 13.6%; Score 167.2; DB 20; Length 193;  
Best Local Similarity 98.3%; Pred. No. 4.5e-35;  
Matches 169; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 385 CTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTC 444  
DB 193 CTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTC 134  
QY 445 CTCGAGCAACACCGCGCGCTTCTACTGTGACTCCCGCACTCCGCCAAGAAATCCG 504  
|||||

DB 133 CTCCAGCAACACCCGCGCCCTTCTAACTGTGACTCTCCCGCACTCCCCAAAAGAAATCCG 74  
QY 505 AAAAACCAAAAGAAACACCGCGCTTACCTGGTGGCGAGAGCCTATCCCCA 556  
|||||  
DB 73 AAAAACCAAAAGAAACACCGCGCTTACCTGGTGGCGAGAAAGGTATCTCGA 22

RESULT 6  
Q77534  
ID Q77534 standard; DNA; 297 BP.  
XX  
AC Q77534:  
DT  
DE 23-SEP-1994 (first entry)  
DE Human genome fragment.  
XX  
KW Brain; placenta; bone marrow; genetic analysis; gene mapping;  
KW detection; homology; human; adrenal tissue; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO9401548-A.  
XX  
PD 20-JAN-1994.  
XX  
PF 13-JUL-1993; 93WO-GB01467.  
XX  
PR 13-JUL-1992; 92GB-0014857.  
XX  
PA (MEDI-) MEDICAL RES COUNCIL.  
XX  
PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;  
PI Sibson DR, Starkey M;  
XX  
DR WPI; 1994-035056/04.  
XX New nucleic acid fragment encoding gene products - can be used  
PT for genetic analysis and mapping  
XX  
PS Claim 1: Page 575-576; 616pp; English.  
XX Human nucleic acid fragments, isolated from brain, adrenal tissue,  
CC the placenta or bone marrow comprise any of: (A) a sequence  
CC selected from (076401-077613), (B) an allelic variation of a  
CC sequence as described in (A), or (C) a sequence complementary  
CC to (A) or (B).  
CC Preferred sequences exhibit no more than 90% homology to a human  
CC sequence known per se.  
XX  
SQ Sequence 297 BP; 66 A; 117 C; 61 G; 52 T; 1 other;

Query Match 12.3%; Score 151; DB 15; Length 297;  
Best Local Similarity 76.0%; Pred. No. 1e-30;  
Matches 212; Conservative 0; Mismatches 65; Indels 2; Gaps 2;

QY 1 ATGTGTCACTCTCGACCTGCCACCCGACCATGACCATCTGTGAGAGCCCGCCGCC 60  
DB 16 atggaacactctcgcgctgcccactcccccatgaccacccctacagggccctacccccagcc 75  
QY 61 CCCTCCACCATCCCGGGGCTCCGGGCTCTGAGATCTTTCACCTTCGAGCCT 120  
DB 76 cctcaggtatctccagccacaaata-gctgcgcctgagatctccactcagaccct 134  
QY 121 CTCGCGAGCCCGCAGCGGCCCTGTGCGGGCGCCCGCAGCGCTCTGCGGGGACCGGAAG 180  
DB 135 ctccgaagccccacagcgccccctgcgggacccccagcttc-ctcgagggcaccacaaag 193  
QY 181 CGCAGCGCGAGGGTTCTCTACCTTCGAGTGTGTCGGCGCGCAGCTGCGAGGACCG 240  
DB 194 cgcagttccagggaatctctaccctagtagtgcggcgccagttgcagagttaaggagacca 253  
|||||



[illegible]

RESULT	9
A02488/c	
ID	A02488 standard; cDNA; 1218 BP.
XX	
XX	
AC	A02488;
XX	
XX	
DT	19-MAY-2000 (first entry)
XX	
DE	Human colon cancer cell line polynucleotide sequence SEQ ID NO:2479
XX	
KW	Human; colon cancer; tumour; diagnosis; gene expression product;
KW	probe; detection; cancerous state; metastasis; identification;
KW	breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW	oestrogen receptor-negative breast cancer; lung cancer; ss.

OS	Homo sapiens.	
XX		
PN	WO958675-A2.	
XX		
PD	18-NOV-1999.	
XX		
PF	13-MAY-1999;	99WO-US10602.
XX		
PR	14-MAY-1998;	98US-0085426.
PR	15-MAY-1998;	98US-0085537.
PR	15-MAY-1998;	98US-0085696.
PR	21-OCT-1998;	98US-0105234.
PR	21-OCT-1998;	98US-0105877.
XX		
PA	(CHIR ) CHIRON CORP.	
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;	
PI	Reinhard G, Giese K, Randazzo F, Kennedy CC, Pot D, Kassam A;	
PI	Lamson G, Drmanac R, Krkenjakov R, Dickson M, Drmanac S, Labat I;	
PI	Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;	
XX		
DR	WPI: 2000-126369/11.	
XX		
PT	Polynucleotide library used to determine cancerous states of mammalian	
PT	cells -	
XX		
PS	Claim 1: Page 995-996; 1097pp: English.	
XX		
CC	A00010 to A02716 represent polynucleotides isolated from CDNA libraries	

constructed from human colon cancer cell lines. The present invention also describes a method of detecting differentially expressed genes correlated with a cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample derived from a cell suspected of being cancerous, where detection of the differentially expressed gene product is correlated with a cancerous state of the cell from which the test sample was derived. The polynucleotide sequences can be used in a method for detecting differentially expressed genes correlated with a cancerous state of a mammalian cell. The polynucleotides can also be used as probes for detecting and mapping related genes. They can be used in diagnosis and prognosis of diseases and disorders (e.g. identification of pre-metastatic or metastatic cancerous states, stages of cancer, or responsiveness of cancer to therapy). This is particularly for breast cancer, oestrogen receptor-positive breast cancer, estrogen receptor-negative breast cancer, lung cancer, and colon cancer.

Sequence 1218 BP: 9 A; 31 C; 494 G; 37 T; 647 other;

[illegible]

RESULT	10
Z17263	
ID	Z17263 standard; cDNA; 1337 BP.
XX	
XX	
AC	Z17263;
XX	
XX	
DT	12-OCT-1999 (first entry)
XX	
XX	
DE	Human gene expression product cDNA sequence SEQ ID NO:4735.
XX	
KW	Human; gene; gene expression product; diagnosis; therapy; probe;
KW	detection; mapping; tissue typing; profiling; forensic; cancer;
KW	genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX	
OS	Homo sapiens.
XX	
XX	





FH	Key	Location/Qualifiers
FT	CDS	31..4014
FT		/tag= a
FT		/product= "Human CNAP-2 protein"
FT		/note= "Shares 24% identity to Aquifex pyrophilus esterase 28LC"
FT	sig_peptide	31..132
FT		/tag= b
FT	mat_peptide	133..4011
FT		/tag= c
FT		/product= "Mature CNAP-2 protein"
FT	misc_binding	136..165
FT		/tag= d
FT		/note_moiety= "Primer or Probe"
FT		/note= "Useful for amplification or hybridisation techniques"
XX		
PX	WO200014248-A1.	
PN		
PD	16-MAR-2000.	
XX		
PF	03-SEP-1999;	99WO-US20287.
XX		
PR	04-SEP-1998;	98US-0148904.
XX		
PA	(INCY-)	INCYTE PHARM INC.
XX		
PI	Hillman JL, Yue H, Guegler KJ, Corley NC, Patterson C, Tang YT;	
XX		
DR	WPI: 2000-256994/22.	
DR	P-PSDB; Y70474.	
XX		
PT	Isolated cyclic nucleotide associated proteins useful for preventing, diagnosing and treating cell proliferative, autoimmune/inflammatory, neurological, vision, reproductive and smooth muscle disorders -	
PT	Claim 9: Page 70-71: 78pp; English.	
XX		
CC	The present sequence is the cDNA encoding human cyclic nucleotide associated protein-2 (CNAP-2), identified in Incyte clone 3149674, that is isolated from ADREN004 CDNA library. It is expressed in nervous, reproductive, cardiovascular and haematopoietic/immune tissues. CNAP sequences may be used for prevention, treatment and diagnosis of diseases associated with altered CNAP expression such as, cell proliferative disorders (e.g. arteriosclerosis, cirrhosis, leukaemia, lymphoma and cancer of the breast, prostate, lung and brain), autoimmune/inflammatory disorders (e.g. asthma, anaemia, diabetes mellitus, multiple sclerosis and psoriasis), neurological disorders (e.g. epilepsy, Alzheimer's/parkinson's disease and strokes), vision disorders (e.g. conjunctivitis, glaucoma, cataracts and retinitis pigmentosa), reproductive disorders (e.g. infertility, uterine fibroids, ectopic pregnancies and impotence) and smooth muscle disorders (e.g. angina, anaphylactic shock, Kearns-Sayre syndrome and hypertension). It can also be used for gene therapy.	
XX		
SQ	Sequence 4228 BP; 793 A; 1365 C; 1324 G; 746 T; 0 other;	
Query Watch            3.9%; Score 47.8; DB 21; Length 4228;		
Best Local Similarity 48.2%; Pred. No. 0.0063;		
Matches 202; Conservative 0; Mismatches 207; Indels 10; Gaps 2		
QY	13 CGCAGCTGCCACCCCACCATGACCATCTCTGCAGCGCCCGCCGGCCCCCTCCACCATC 72	
Db		
Db	907 clcttcagccagcagatccagcccctcgctgttccccagccgcggccclcccaactcgc 966	
QY	73 CCGGACCCCGCGGGGTTCGGTCTTGAGATCTTACCTTCGACCTCTCTCCCG----- 126	
Db		
QY	127 --GAGCCGACGACGCCCTCTGCGGGCGCCCGCCAGCGCTCTCGCGGCCCGAAGCGCA 184	
Db		
Db	1027 gagaccacagcaggccaccgcgatcccaacgggggcccgctgctgacctacagggac 1086	





Search completed: January 30, 2001, 21:29:29  
Job time: 19255 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 23:11:29 ; Search time 64.84 Seconds  
(without alignments)  
155.458 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156

Perfect score: 442  
Sequence: 1 VRRQLPVEPNPAKRLFL...EPSDYALDLSTFLQHPAAF 86

Scoring table: BLOSUM62 %  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SPTREMBL15:\*

- 1: sp-archaea:\*
- 2: sp-bacteria:\*
- 3: sp-fungi:\*
- 4: sp-human:\*
- 5: sp-invertebrate:\*
- 6: sp-mammal:\*
- 7: sp-mhc:\*
- 8: sp-organelle:\*
- 9: sp-phage:\*
- 10: sp-plant:\*
- 11: sp-rodent:\*
- 12: sp-virus:\*
- 13: sp-vertebrate:\*
- 14: sp-unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	442	100.0	193	4	075353
2	84.5	19.1	637	4	094878
3	81.5	18.4	102	5	0972D3
4	81.5	18.4	452	5	03UAS5
5	81.5	18.4	468	5	09XY65
6	80.5	18.2	302	12	09QM69
7	78	17.6	4957	4	014687
8	78	17.6	5262	4	014686
9	77	17.4	1568	4	095785
10	77	17.4	2424	5	09VZ48
11	76	17.2	148	2	09J2R7
12	76	17.2	801	5	09A3K6
13	76	17.2	1677	5	000805
14	74	16.7	114	12	09YK08
15	74	16.7	955	11	088287
16	74	16.7	1040	5	09VH04
17	74	16.7	1561	11	088286
18	73.5	16.6	925	12	069138
19	73	16.5	1029	12	09YMX0

20	72	16.3	559	5	09VQ23	09vq23 drosophila
21	72	16.3	852	4	09NP71	09np71 homo sapien
22	71	16.1	359	5	09TYW7	09tyw7 caenorhabdi
23	71	16.1	1882	3	09P898	09p898 podospora a
24	70.5	16.0	244	12	069126	069126 human herpe
25	70.5	16.0	1900	4	015022	015022 homo sapien
26	70.5	16.0	2605	4	09UK25	09uk25 homo sapien
27	70.5	16.0	2715	4	09UMN6	09umn6 homo sapien
28	70	15.8	148	2	09JUV2	09juv2 neisseria m
29	70	15.8	1233	2	09RI32	09ri32 streptomyce
30	70	15.8	2971	4	09Y5L9	09y5l9 homo sapien
31	69.5	15.7	277	4	09XN6	09xn6 homo sapien
32	69.5	15.7	562	4	09NVJ4	09nvj4 homo sapien
33	69.5	15.7	745	4	09P2N8	09p2n8 homo sapien
34	69.5	15.7	757	6	09N015	09n015 macaca fasc
35	69.5	15.7	831	4	09UMS6	09ums6 homo sapien
36	69.5	15.7	1887	3	012533	012533 saccharomyc
37	69	15.6	418	5	018503	018503 caenorhabdi
38	69	15.6	646	4	09ULX6	09ulx6 homo sapien
39	69	15.6	646	4	09UGM0	09ugm0 homo sapien
40	69	15.6	1454	5	010463	010463 caenorhabdi
41	69	15.6	2157	11	0921R1	0921r1 mus musculu
42	69	15.6	2187	11	P70670	P70670 mus musculu
43	68.5	15.5	701	4	09NXX8	09nxx8 homo sapien
44	68.5	15.5	734	4	09NSE6	09nse6 homo sapien
45	68.5	15.5	821	4	09UII8	09uii8 homo sapien

ALIGNMENTS

RESULT 1  
075353 PRELIMINARY; PRT; 193 AA.  
AC 075353;  
DT 01-NOV-1998 (TremBLrel. 08, Created)  
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)  
DT 01-MAY-1999 (TremBLrel. 10, Last annotation update)  
DE ANTI-DEATH PROTEIN.  
CN IEX-1L.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98369175; PubMed=9703517;  
RA Wu M.X., Ao Z., Prasad K.V.S., Wu R., Schlossman S.F.;  
RT "IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival."  
RL Science 281:998-1001(1998).  
DR EMBL; AF039067; AAC32558.1; -;  
DR EMBL; AF071596; AAC72344.1; -;  
SQ SEQUENCE 193 AA; 21028 MW; 7927D9D3FFBC7C57 CRC64;

Query Match 100.0%; Score 442; DR 4; Length 193;  
Best Local Similarity 100.0%; Pred. No. 8.7e-40;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VRRQLPVEPNPAKRLFL...EPSDYALDLSTFLQHPAAF 86  
Db 108 VRRQLPVEPNPAKRLFL...EPSDYALDLSTFLQHPAAF 193  
Qy 61 PFNLTSPSDYALDLSTFLQHPAAF 86  
Db 168 PFNLTSPSDYALDLSTFLQHPAAF 193

RESULT 2  
094878 PRELIMINARY; PRT; 637 AA.  
ID 094878  
AC 094878;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
 DE KIAA0781 PROTEIN (FRAGMENT).  
 GN KIAA0781.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euleleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=99087487; PubMed=9872452;  
 RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,  
 RA Kotani H., Nomura N., Ohata O.,  
 RT "Prediction of the coding sequences of unidentified human genes. XI.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro."  
 RL DNA Res. 5:277-286(1998).  
 DR EMBL; AB018324; BAA34501.1;  
 DR INTERPRO; IPR000449;  
 DR INTERPRO; IPR002965;  
 DR PRINTS; PRO1217; PRICHEXTENSN.  
 FT NON\_TER  
 FT 1  
 SQ SEQUENCE 637 AA; 70516 MW; DEBFB005BA9B9CDD CRC64;

Query Match 19.1%; Score 84.5; DB 4; Length 637;  
 Best Local Similarity 33.3%; Pred. No. 0.41;  
 Matches 30; Conservative 12; Mismatches 27; Indels 21; Gaps 5;  
 QY 7 VEENPAKRLLEL---LTIVFCQILMAEEGV-----APLPEDAPNAASLAP----- 52  
 DB 423 LGEHRLQRLFLQKQSLQAVFNQMQIAESSYPOPSQOLPLPROETPTPPSQOAPPFSLT 482  
 QY 53 TPVSPVLEPNLTSEPSDYALDLSTFLQOH 82  
 DB 483 QPLSPVLEP---SSEOMQV---SPFLSQY 505

RESULT 3  
 ID Q9T2D3 PRELIMINARY; PRT; 102 AA.  
 AC Q9T2D3;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE R160.5 PROTEIN.  
 GN R160.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,  
 RA Smaison N., Smith A., Sonhammer E., Staden K., Sulston J.,  
 RA Therry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Wellstock L., Wilkinson-Sproat J., Wohlman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans."  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;

RA Nelson J., Gattung S.;  
 RT "The sequence of C. elegans cosmid R160.";  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF099001; AAC68734.1;  
 SQ SEQUENCE 102 AA; 11426 MW; AFD366C3932EB26C CRC64;

Query Match 18.4%; Score 81.5; DB 5; Length 102;  
 Best Local Similarity 24.7%; Pred. No. 0.13;  
 Matches 19; Conservative 19; Mismatches 28; Indels 11; Gaps 3;  
 QY 11 NPAKRLLEL---LTIVFCQILMAEEGVAPLPEDAPNAASLAPVPSPVLEPFN-- 63  
 DB 7 NPAIQFAFVVILVCAVLCIMLIERADGFCPLPPEELRIVNTIP-PRTPIAAPVNR 65  
 QY 64 ---LTSEPSDYALDLST 77  
 DB 66 QVVIKIDPTESDVMDS 82

RESULT 4  
 ID Q9UAS5 PRELIMINARY; PRT; 452 AA.  
 AC Q9UAS5;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE DEVELOPMENTAL-SPECIFIC PROTEIN CONZAB (FRAGMENT).  
 GN CONZAB.  
 OS Euplotes crassus.  
 OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotida; Euplotidae;  
 OC Euplotes.  
 OX NCBI\_TaxID=5936;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ST9, ST11;  
 RX MEDLINE=97206572; PubMed=9172827;  
 RA Ling Z., Ghosh S., Jacobs M.E., Klobutcher L.A.;  
 RT "Conjugation-specific genes in the ciliate Euplotes crassus: gene  
 RT expression from the old macronucleus."  
 RL J. Eukaryot. Microbiol. 44:1-11(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ST9, ST11;  
 RA Jacobs M.E., Ling Z., Klobutcher L.A.;  
 RT "conzab encodes a novel and abundant protein targeted to the  
 RT developing macronucleus in Euplotes crassus."  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF116196; AAD29625.1;  
 FT NON\_TER  
 FT 1  
 SQ SEQUENCE 452 AA; 49763 MW; A5C13C49ECA998EA CRC64;

Query Match 18.4%; Score 81.5; DB 5; Length 452;  
 Best Local Similarity 34.5%; Pred. No. 0.61;  
 Matches 30; Conservative 10; Mismatches 30; Indels 17; Gaps 5;  
 QY 3 RQLPVEEP----NPAKRLLEL---LTIVFCQILMAEEGVAPLPEDAPNAASLAP----P 54  
 DB 316 RQSTPLPLILENPAKEIASLSKMI---QETAIESIPTQPP-----VSSIQPSAPVHP 367  
 QY 55 VSPVLEPFN-LTSEPSDYALDLSTFLQ 80  
 DB 368 SAPSPSSTLASEPRDIIIPDPSAALK 394

RESULT 5  
 Q9XY65

ID Q9XY65 PRELIMINARY; PRT: 468 AA.  
AC Q9XY65;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE CONZAB PROTEIN.  
GN CONZAB.  
OS Euplotes crassus.  
OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotida; Euplotidae;  
OC Euplotes.  
OX NCBI\_TaxID=5936;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G1;  
RX MEDLINE=97206572; PubMed=9172827;  
RA Ling Z., Ghosh S., Jacobs M.E., Klobutcher L.A.;  
RT "Conjugation-specific genes in the ciliate Euplotes crassus: gene  
RT expression from the old macronucleus."  
RL J. Eukaryot. Microbiol. 44:1-11(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G1;  
RX MEDLINE=98391748; PubMed=9722644;  
RA Klobutcher L.A., Gyax S.E., Podoloff J.D., Vermeesch J.R.,  
RA Price C.M., Tebeau C.M., Jahn C.L.;  
RT "Conserved DNA sequences adjacent to chromosome fragmentation and  
RT telomere addition sites in Euplotes crassus."  
RL Nucleic Acids Res. 26:4230-4240(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G1;  
RA Ling Z., Klobutcher L.A.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF061334; AAD31724.1;  
SQ SEQUENCE 468 AA; 51457 MW; 70996125317C592A CRC64;

Query Match 18.4%; Score 81.5; DB 5; Length 468;  
Best Local Similarity 34.5%; Pred. No. 0.63;  
Matches 30; Conservative 10; Mismatches 30; Indels 17; Gaps 5;

QY 3 ROLPVEEP---NPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAA--P 54  
Db 332 RSPPLPLDLENPAKETASSLSKMI---QETAIESIPTQPQ-----VSSIQPSAPVHP 383  
QY 55 VSPVLEPFN-LTSEPSDYALDLSFTLQ 80  
Db 384 SAPSLEPSSTLASEPRDIIPDPSAALK 410

RESULT 6  
Q9QM69 PRELIMINARY; PRT: 302 AA.  
AC Q9QM69;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE 33 KDA PROTEIN.  
OS fowl adenovirus 8.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.  
OX NCBI\_TaxID=66295;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ojtic D., Nagy E.;  
RT "The DNA sequence of fowl adenovirus 8."  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF083975; AAD50348.1;  
SQ SEQUENCE 302 AA; 34763 MW; BF34DF42FF026CCC CRC64;

Query Match 18.2%; Score 80.5; DB 12; Length 302;  
Best Local Similarity 26.4%; Pred. No. 0.51;  
Matches 24; Conservative 26; Mismatches 30; Indels 11; Gaps 4;

QY 2 BRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLP-----PEDAPNAA--SLAPT 53  
Db 171 RKLLDLTDESDRRKSPFFMF-LVSDDAVIDEEPVPPI.PKRRRRRANKEPNASPTQLPE 229  
QY 54 PVSVPLEPFNLTSEPSDYALDLSFTLQHPA 84  
Db 230 PVS PAVS--DLKAEILNLLVEIESFVRKNPS 258  
RESULT 7  
O14687 PRELIMINARY; PRT: 4957 AA.  
AC O14687;  
DT 01-JAN-1998 (TReMBLrel. 05, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
DE ALR.  
GN ALR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97388474; PubMed=9247308;  
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,  
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,  
RA Canani E.;  
RT "Structure and expression pattern of human ALR, a novel gene with  
RT strong homology to ALL-1 involved in acute leukemia and to Drosophila  
RT trithorax."  
RL Oncogene 15:549-560(1997).  
DR EMBL: AF010404; AAC51735.1;  
DR INTERPRO: IPR001214;  
DR INTERPRO: IPR001822;  
DR INTERPRO: IPR001965;  
DR INTERPRO: IPR002965;  
DR PFAM: PF00628; PHD; 3.  
DR PFAM: PF00856; SET; 1.  
DR PRINTS: PR01217; PRICHEXTENS.  
DR PROSITE: PS00398; RECOMBINASE\_2; UNKNOWN\_1.  
SQ SEQUENCE 4957 AA; 531848 MW; 1026562E1419CE8D CRC64;

Query Match 17.6%; Score 78; DB 4; Length 4957;  
Best Local Similarity 30.4%; Pred. No. 17;  
Matches 21; Conservative 12; Mismatches 28; Indels 8; Gaps 2;

QY 1 VRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAA--SLAPTVPSPVLE 60  
Db 4299 LKQESPAPEPTQHRVTYVNSNLDVRQL-----SAPPEPSPPSPSPSPSPTE 4351  
QY 61 PF-NLTSEP 68  
Db 4352 PLVELPTEP 4360

RESULT 8  
O14686 PRELIMINARY; PRT: 5262 AA.  
AC O14686;  
DT 01-JAN-1998 (TReMBLrel. 05, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
DE ALR.  
GN ALR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=97388474; PubMed=9247308;  
 RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,  
 RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,  
 RA Canaan E.  
 RT "Structure and expression pattern of human ALR, a novel gene with  
 RT strong homology to ALL-1 involved in acute leukemia and to Drosophila  
 RT trithorax";  
 RL Oncogene 15:549-560(1997).  
 DR EMBL: AF010403; AAC51734.1; -  
 DR INTERPRO: IPR001214; -  
 DR INTERPRO: IPR001822; -  
 DR INTERPRO: IPR001841; -  
 DR INTERPRO: IPR001965; -  
 DR PFAM: PF00628; PHD: 5.  
 DR PFAM: PF00856; SET: 1.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 DR PROSITE: PS00398; RECOMBINASES\_2; UNKNOWN\_1.  
 SQ SEQUENCE 5262 AA; 564181 MW; 26B7C74CAD17B44 CRC64;

Query Match 17.6%; Score 78; DB 4; Length 5262;  
 Best Local Similarity 30.4%; Pred No. 18;  
 Matches 21; Conservative 12; Mismatches 28; Indels 8; Gaps 2;

QY 1 VRQLPVEEDNPAKRLFLLLTIVFCOILMAEEGVAPLPDPDAPNAASLAPTPVSPVLE 60  
 Db 4604 LQOESAPPEPQHRYTYNVNSLDVRQL-----SAPPEEPSPPSPPLAPSPAPTE 4656  
 QY 61 PF-NLTSEP 68  
 Db 4657 PLVELDPT 4655

RESULT 9  
 ID O95785 PRELIMINARY; PRT: 1568 AA.  
 AC O95785;  
 DT 01-MAY-1999 (TRENBLrel. 10, Created)  
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
 DE HUMAN HOMOLOG OF MUS MUSCULUS WIZL PROTEIN (HUMAN HOMOLOG OF MUS  
 DE MUSCULUS WIZS PROTEIN) (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,  
 RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stilwagen S.,  
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Barnes J.,  
 RA Danganan L., Erlar A., Christensen M., Georgescu A., Avila J., Liu S.,  
 RA Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,  
 RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,  
 RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,  
 RA Olsen A.S., Carrano A.V.;  
 RA "Sequence analysis of a 3.5 Mb region in 19p13.1 between OLFR and  
 RT JUND";  
 RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,  
 RA Burkhardt-Schultz K., Gordon L., Dias J., Brower A., Ramirez M.,  
 RA Stilwagen S., Phan H., Velasco N., Do L., Regala W., Terry A.,  
 RA Barnes J., Danganan L., Erlar A., Christensen M., Georgescu A.,  
 RA Avila J., Liu S., Attix C., Andreise T., Trankheim M.,  
 RA Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R.,  
 RA Thomas P., Quan G., Kronmiller B., Arellano A., Sanders C., Ow D.,  
 RA Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.;  
 RA "Sequence analysis of an ~1.5 Mb contig in 19p13.1 between OLFR and  
 RT D19S885";  
 RT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC007059; AAD19818.1; -  
 DR EMBL: AC006128; AAC97985.1; -  
 DR EMBL: AC007059; AAD19817.1; -  
 DR INTERPRO: IPR000822; -  
 DR PFAM: PF00096; ZF-C2H2; 10.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2; 8.  
 KW Zinc-finger; Metal-binding; DNA-binding.  
 FT NON\_TER 1  
 SQ SEQUENCE 1568 AA; 170077 MW; 74F0D34D45F565D0 CRC64;

Query Match 17.4%; Score 77; DB 4; Length 1568;  
 Best Local Similarity 40.4%; Pred No. 6.6;  
 Matches 19; Conservative 3; Mismatches 17; Indels 8; Gaps 1;

QY 30 MAEGVPAPLPDPDAPNAASLAPTPVSPVLE-----PFNLTSEP 68  
 Db 908 LALAGSPTPKPNPDKSPQSLSPSPASPKAQMPOSEDEGLPLNLTSGP 954

RESULT 10  
 ID O9V248 PRELIMINARY; PRT: 2424 AA.  
 AC O9V248;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DE 01-OCT-2000 (TRENBLrel. 15, Last annotation update)  
 DE CG2174 PROTEIN.  
 GN CG2174.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]

RA SEQUENCE FROM N.A.  
 RP STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Abayani A., An H.-J., Andrews-Bianknoch C., Baldwin D.,  
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Besley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Flosser C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.M., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
 RA Palazzolo M., Plittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert B.C., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003484; AAF47980.1;
DR HSSP: P06799; 1MND.
DR FLYBASE: FBgn0030252; CG2174.
DR INTERPRO: IPR000048;
DR INTERPRO: IPR000857;
DR INTERPRO: IPR001609;
DR PFAM: PF00063; myosin_head; 1.
DR PFAM: PF00612; IQ; 3.
DR PFAM: PF00784; MYTH4; 1.
DR PRINTS: PROU193; MYOSINHEAVY.
SQ SEQUENCE 2424 AA; 267616 MW; 8ADG2AA33F9AA5D0 CRC64;

Query Match 17.4%; Score 77; DB 5; Length 2424;
Best Local Similarity 36.7%; Pred. No. 10;
Matches 22; Conservative 7; Mismatches 29; Indels 2; Gaps 2;

QY 28 ILMAEGVPAPLPEDAPNAASLAPTPVSPLEPNLTSEPS-DYALDLSTFLOQHPAF 86
Db 1950 LLPPAPSVAPPPIRPPSMAPPAPP-AQSPPTARSPEDPVRTSSQVVKHVPAP 2008

RESULT 11
ID Q9JZR7 PRELIMINARY; PRT; 148 AA.
AC Q9JZR7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ACYL COA THIOESTER HYDROLASE FAMILY PROTEIN.
GN NMB0925.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi C., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
DR EMBL: AE002444; AAF41333.1;
DR TIGR: NMB0925;
KW Hydrolase.
SQ SEQUENCE 148 AA; 16332 MW; 28E9B40D502D80B4 CRC64;

Query Match 17.2%; Score 76; DB 2; Length 148;
Best Local Similarity 40.5%; Pred. No. 0.75;
Matches 15; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY 18 FLLLTIVFCQILMAEGVPAPLPEDAPNAASLAPTP 54
Db 112 YLVTEAVFTYVAIDAEGRNPRIPKXGNPKLAGLLPT 148

RESULT 12
Q9W3K6
ID Q9W3K6 PRELIMINARY; PRT; 801 AA.
AC Q9W3K6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

```

```

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG2258.
GN PROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flossler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paeteb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003442; AAF46320.1;
DR HSSP: P06241; 1SHF.
DR FLYBASE: FBgn0029997; CG2258.
DR INTERPRO: IPR001452;
DR PFAM: PF00018; SH3; 1.
DR PROSITE: PS50002; SH3; 1.
SQ SEQUENCE 801 AA; 89233 MW; A0F5C824663DC391 CRC64;

Query Match 17.2%; Score 76; DB 5; Length 801;
Best Local Similarity 33.9%; Pred. No. 4.2;
Matches 21; Conservative 8; Mismatches 23; Indels 10; Gaps 2;

QY 26 COILMAEGVPAPLPEDA--PNAASLAPTPVS-----PYLEPNLTSEPSYALDL 75
Db 206 CTIQTQQOQPLPAPPAPAPSPSAASATPTAASTSKAATGQDVAQNSITSEPTTTT 265

QY 76 ST 77
Db 266 TT 267

RESULT 13

```

000805  
ID 000805 PRELIMINARY: PRT: 1677 AA.  
AC 000805;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DE 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE P-GLYCOPROTEIN E.  
OS Leishmania tropica.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=5666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LRC-L39;  
RA Lafuente E., Castanys S., Camarro F.;  
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U55381; AAB51191.1;  
DR HSSP; P13569; 1NBD.  
DR INTERPRO; IPR000504;  
DR INTERPRO; IPR001140;  
DR INTERPRO; IPR001617;  
DR INTERPRO; IPR001865;  
DR PFAM; PF00005; ABC\_tran; 2.  
DR PFAM; PF00664; ABC\_membrane; 2.  
DR PROSITE; PS00030; RNP\_1; UNKNOWN\_1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
DR PROSITE; PS00962; RIBOSOMAL\_S2\_1; UNKNOWN\_1.  
SQ SEQUENCE 1677 AA; 183006 MW; C49208921B71563A CRC64;

Query Match 17.2%; Score 76; DB 5; Length 1677;  
Best Local Similarity 33.9%; Pred. No. 9;  
Matches 20; Conservative 6; Mismatches 15; Indels 18; Gaps 2;

QY 37 APLP-----PEDANNAASLAPTPVSPVLEFPN-----LTSEPSDYALDLST 77  
||| | : : : | | | | | : : : | |  
Db 698 APLPEAEAPLPDPQSSSSAAPPAGVTEPLNAGKSHGTGHGDAASSEPLSSSAQKST 756

RESULT 14  
QYQY08 PRELIMINARY: PRT: 114 AA.  
AC QYQY08;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
DE LATE 33 K PROTEIN HOMOLOG.  
OS fowl adenovirus 8.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.  
OX NCBI\_TaxID=66295;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC A-2A;  
RX MEDLINE=98451338; PubMed=9780058;  
RA Cao J.X., Kreil P.J., Nagy E.;  
RT "Sequence and transcriptional analysis of terminal regions of the fowl  
adenovirus type 8 genome."  
RL J. Gen. Virol. 79:2507-2516(1998).  
DR EMBL; AF021254; AAC71673.1;  
SQ SEQUENCE 114 AA; 13266 MW; 866DEEC6636EABA6 CRC64;

Query Match 16.7%; Score 74; DB 12; Length 114;  
Best Local Similarity 31.1%; Pred. No. 0.94;  
Matches 19; Conservative 17; Mismatches 15; Indels 10; Gaps 3;

QY 32 EGVVPAPLP-----PEDANNAASLAPTPVSPVLEFPNLTSEPSDYALDLSTFLOQHP 83  
||| | | : : : | | | | : : : | | : : : | |  
Db 12 EEPVPPPLPKRRRRANKKEPNASSETQLPEVSPAVS--DLKAEILNLLVEIESFVRKNP 69  
QY 84 A 84  
Db 70 S 70

RESULT 15  
Q88287 PRELIMINARY: PRT: 955 AA.  
ID 088287  
AC 088287;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE WIZ.  
GN WIZ.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RA Matsumoto K., Ishii N., Yoshida S., Shiosaka S., Wanaka A.,  
RA Tohyama M.;  
RT "Molecular Cloning and Distinct Developmental Expression Pattern of  
RT Spliced Forms of A Novel Zinc Finger Gene wiz in The Cerebellum."  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB012266; BAA32791.1;  
DR MGD; MGI:1332638; Wiz.  
DR INTERPRO; IPR000822;  
DR PRINTS; PR00048; ZINC\_FINGER.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2; 5.  
KW Zinc-finger; Metal-binding; DNA-binding.  
SQ SEQUENCE 955 AA; 102777 MW; D4DD03A02BECCAAC CRC64;  
Query Match 16.7%; Score 74; DB 11; Length 955;  
Best Local Similarity 41.9%; Pred. No. 8.3;  
Matches 18; Conservative 2; Mismatches 15; Indels 8; Gaps 1;  
QY 34 GVPAPLPEDAPNNAASLAPTPVSPVLE-----PFNLTSSEP 68  
| | | | | : : : | | | : | | | |  
Db 300 GSPTPKNPEDKSPQLSLSRPRTSPRAOCPQSEDEGLNLATSGP 342

Search completed: January 30, 2001, 23:11:30  
Job time: 3995 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 21:29:34 ; Search time 57.93 seconds  
(without alignments)  
401.549 Million cell updates/sec

Title: US-08-799-910-10

Perfect score: 823

Sequence: 1 MCHSRCHPTMTILOAPTPA.....EPDVALDSTFLOQHPAAF 156

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 912318 seqs, 149113690 residues

Total number of hits satisfying chosen parameters: .912318

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/paa/US080\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*
- 9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*
- 10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*
- 11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*
- 12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*
- 13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*
- 14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*
- 15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*
- 16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*
- 17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*
- 18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*
- 19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*
- 20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*
- 21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep.\*
- 22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*
- 23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep.\*
- 24: /cgn2\_6/ptodata/2/paa/US100\_COMB.pep.\*
- 25: /cgn2\_6/ptodata/2/paa/US101\_COMB.pep.\*
- 26: /cgn2\_6/ptodata/2/paa/US102\_COMB.pep.\*
- 27: /cgn2\_6/ptodata/2/paa/US103\_COMB.pep.\*
- 28: /cgn2\_6/ptodata/2/paa/US104\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
1	823	100.0	156 11	US-08-799-910-10 Sequence 10, Appl
2	823	100.0	156 12	US-08-825-486-10 Sequence 10, Appl
3	823	100.0	156 12	US-08-826-248-10 Sequence 10, Appl
4	823	100.0	156 12	US-08-870-434-11 Sequence 11, Appl
5	823	100.0	156 13	US-08-925-588-10 Sequence 10, Appl

6	823	100.0	156	17	US-09-372-044-10	Sequence 10, Appl
7	794	96.5	151	22	US-60-197-873-17386	Sequence 17386, A
8	440	53.5	123	22	US-60-196-718-6102	Sequence 6102, Ap
9	425	51.6	83	1	PCT-US00-15136-168	Sequence 168, App
10	423	51.4	83	1	PCT-US00-15136-96	Sequence 96, Appl
11	422	51.3	83	1	PCT-US00-15136-167	Sequence 167, App
12	381	46.3	89	22	US-60-196-718-4945	Sequence 4945, Ap
13	358	43.5	71	1	PCT-US00-15136-169	Sequence 169, App
14	358	43.5	92	22	US-60-196-718-5439	Sequence 5439, Ap
15	340	41.3	72	1	PCT-US00-15136-170	Sequence 170, App
16	273.5	33.2	56	22	US-60-196-718-3936	Sequence 3936, Ap
17	114.5	13.9	1828	22	US-60-188-986-15	Sequence 15, Appl
18	114.5	13.9	2012	22	US-60-229-524-80	Sequence 80, Appl
19	114.5	13.9	2972	19	US-09-579-181-2	Sequence 2, Appl1
20	114.5	13.9	3041	28	US-60-258-275-409	Sequence 409, App
21	114.5	13.9	3118	19	US-09-579-181-1	Sequence 1, Appl1
22	112.5	13.7	649	22	US-60-191-637-28141	Sequence 28141, A
23	112.5	13.7	649	22	US-60-191-681-22750	Sequence 22750, A
24	112.5	13.7	1255	22	US-60-167-217-13778	Sequence 13778, A
25	112.5	13.7	1255	22	US-60-173-464-11201	Sequence 11201, A
26	112	13.6	747	18	US-09-417-507-31296	Sequence 31296, A
27	110	13.4	590	1	PCT-US99-22853B-2801	Sequence 2801, Ap
28	109	13.2	8991	11	US-08-714-741-32	Sequence 32, Appl
29	108	13.1	271	16	US-09-248-796-21050	Sequence 21050, A
30	108	13.1	271	22	US-60-096-409-21050	Sequence 21050, A
31	107.5	13.1	1013	22	US-60-191-637-39541	Sequence 39541, A
32	107.5	13.1	1013	22	US-60-191-681-30574	Sequence 30574, A
33	107	13.0	206	9	US-08-529-055-54	Sequence 54, Appl
34	106	12.9	1110	22	US-60-243-468-898	Sequence 898, App
35	105.5	12.8	1013	22	US-60-173-464-30071	Sequence 30071, A
36	105	12.8	296	22	US-60-173-464-18776	Sequence 18776, A
37	105	12.8	296	22	US-60-191-637-22907	Sequence 22907, A
38	105	12.8	296	22	US-60-191-681-18056	Sequence 18056, A
39	104	12.6	129	9	US-08-529-055-53	Sequence 53, Appl
40	102	12.4	331	22	US-60-173-464-29642	Sequence 29642, A
41	102	12.4	331	22	US-60-191-637-38878	Sequence 38878, A
42	102	12.4	331	22	US-60-191-681-30133	Sequence 30133, A
43	102	12.4	421	18	US-09-451-320-4244	Sequence 4244, Ap
44	101.5	12.3	193	9	US-08-529-055-49	Sequence 49, Appl
45	101.5	12.3	741	18	US-09-417-507-37506	Sequence 37506, A

#### ALIGNMENTS

RESULT 1  
US-08-799-910-10  
; Sequence 10, Application US/08799910  
; GENERAL INFORMATION:  
; APPLICANT: Falb, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/799,910  
; FILING DATE: 13-FEB-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/011,787  
; FILING DATE: 16-FEB-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-067-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 156 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-799-910-10

Query Match 100.0%; Score 823; DB 11; Length 156;  
Best Local Similarity 100.0%; Pred. No. 3.2e-60;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MCHSRSCPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60  
DB 1 MCHSRSCPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60  
QY 61 RSRRLVPRVVRQLPVEEPNPAKRLLFLLLTIVFCQILMAEGVPAPLPDPEDAPNAASL 120  
DB 61 RSRRLVPRVVRQLPVEEPNPAKRLLFLLLTIVFCQILMAEGVPAPLPDPEDAPNAASL 120  
QY 121 APTVPSPVLEPFLNTSEPSDYALDSTFLOOHPAAF 156  
DB 121 APTVPSPVLEPFLNTSEPSDYALDSTFLOOHPAAF 156

RESULT 2  
US-08-825-486-10  
Sequence 10, Application US/08825486  
GENERAL INFORMATION:  
APPLICANT: Falb, Dean  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/825,486  
FILING DATE: 28-MAR-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/799,910  
FILING DATE: 13-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-077-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:  
LENGTH: 156 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-825-486-10

Query Match 100.0%; Score 823; DB 12; Length 156;  
Best Local Similarity 100.0%; Pred. No. 3.2e-60;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MCHSRSCPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60  
DB 1 MCHSRSCPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60  
QY 61 RSRRLVPRVVRQLPVEEPNPAKRLLFLLLTIVFCQILMAEGVPAPLPDPEDAPNAASL 120  
DB 61 RSRRLVPRVVRQLPVEEPNPAKRLLFLLLTIVFCQILMAEGVPAPLPDPEDAPNAASL 120  
QY 121 APTVPSPVLEPFLNTSEPSDYALDSTFLOOHPAAF 156  
DB 121 APTVPSPVLEPFLNTSEPSDYALDSTFLOOHPAAF 156

RESULT 3  
US-08-826-248-10  
Sequence 10, Application US/08826248  
GENERAL INFORMATION:  
APPLICANT: Falb, Dean  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/826,248  
FILING DATE: 28-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/799,910  
FILING DATE: 13-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/011,787  
FILING DATE: 16-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-079-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 156 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein

```
; FRAGMENT TYPE: internal
US-08-826-248-10

Query Match      100.0%; Score 823; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVPAPLPDPEDAPNAASL 120
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVPAPLPDPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOOHPPAAAF 156
DB 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOOHPPAAAF 156

RESULT 4
US-08-870-434-11
; Sequence 11, Application US/08870434
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870.434
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799.910
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-084
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; MOLECULE TYPE: protein
; US-08-870-434-11

Query Match      100.0%; Score 823; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
```

```
; FRAGMENT TYPE: internal
US-08-925-588-10

Query Match      100.0%; Score 823; DB 13; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVPAPLPDPEDAPNAASL 120
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVPAPLPDPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOOHPPAAAF 156
DB 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOOHPPAAAF 156

RESULT 5
US-08-925-588-10
; Sequence 10, Application US/08925588
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925.588
; FILING DATE: 08-Sep-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799.910
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-08-925-588-10

Query Match      100.0%; Score 823; DB 13; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.2e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60
DB 1 MCHSRCHPTMTILOAPTAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASRGHRK 60

QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVPAPLPDPEDAPNAASL 120
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVPAPLPDPEDAPNAASL 120

QY 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOOHPPAAAF 156
DB 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLOOHPPAAAF 156
```

RESULT 6  
US-09-372-044-10  
; Sequence 10, Application US/09372044A  
; GENERAL INFORMATION:  
; APPLICANT: Dean FALB et al.  
; TITLE OF INVENTION: Compositions and Methods for the  
; TITLE OF INVENTION: Treatment and Diagnosis of Cardiovascular Disease  
; FILE REFERENCE: 7853-152  
; CURRENT APPLICATION NUMBER: US/09/372.044A  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 156  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-372-044-10

Query Match 100.0%; Score 823; DB 17; Length 156;  
Best Local Similarity 100.0%; Pred. No. 3.2e-60;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MCHSRSCPTMTILOAPTAPSTIPGRRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60  
Db 1 MCHSRSCPTMTILOAPTAPSTIPGRRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60  
QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120  
Db 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120  
QY 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLQHPAAF 156  
Db 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLQHPAAF 156

RESULT 7  
US-60-197-873-17386  
; Sequence 17386, Application US/60197873  
; GENERAL INFORMATION:  
; APPLICANT: Bejanin, Stephane  
; APPLICANT: Tanaka, Hiroaki  
; APPLICANT: Dumas Milne Edwards, Jean Baptiste  
; APPLICANT: Jobert, Severin  
; APPLICANT: Giordano, Jean-Yves  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: 81 US1 PRO  
; CURRENT APPLICATION NUMBER: US/60/197,873  
; NUMBER OF SEQ ID NOS: 52153  
; SOFTWARE: Patent.pm  
; SEQ ID NO 17386  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-197-873-17386

Query Match 96.5%; Score 794; DB 22; Length 151;  
Best Local Similarity 100.0%; Pred. No. 7.5e-58;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MCHSRSCPTMTILOAPTAPSTIPGRRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60  
Db 1 MCHSRSCPTMTILOAPTAPSTIPGRRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60  
QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120  
Db 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120  
QY 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLQ 151

Db 121 APTVPSPVLEPFLNTSEPSDYALDLSTFLQ 151  
RESULT 8  
US-60-196-718-6102  
; Sequence 6102, Application US/60196718  
; GENERAL INFORMATION:  
; APPLICANT: Bonazzi, Vivien  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL000456  
; CURRENT APPLICATION NUMBER: US/60/196.718  
; CURRENT FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7494  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 6102  
; LENGTH: 123  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-60-196-718-6102

Query Match 53.5%; Score 440; DB 22; Length 123;  
Best Local Similarity 93.5%; Pred. No. 7.4e-29;  
Matches 86; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 65 VLYPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASLAPT 124  
Db 32 LFLCLPLRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASLAPT 91  
QY 125 VSPVLEPFLNTSEPSDYALDLSTFLQHPAAF 156  
Db 92 VSPVLEPFLNTSEPSDYALDLSTFLQHPAAF 123

RESULT 9  
PCT-US00-15136-168  
; Sequence 168, Application PC/TUS0015136  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: 50 Human Secreted Proteins  
; FILE REFERENCE: PS551PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/15136  
; CURRENT FILING DATE: 2000-06-01  
; EARLIER APPLICATION NUMBER: 60/138,629  
; NUMBER OF SEQ ID NOS: 198  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 168  
; LENGTH: 83  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US00-15136-168

Query Match 51.6%; Score 425; DB 1; Length 83;  
Best Local Similarity 100.0%; Pred. No. 8.2e-28;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 70 VVRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASLAPTVPSPVL 129  
Db 1 VVRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASLAPTVPSPVL 60  
QY 130 EPFLNTSEPSDYALDLSTFLQ 152  
Db 61 EPFLNTSEPSDYALDLSTFLQ 83

RESULT 10  
PCT-US00-15136-96  
; Sequence 96, Application PC/TUS0015136

GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: 50 Human Secreted Proteins  
; FILE REFERENCE: PS551PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/15136  
; CURRENT FILING DATE: 2000-06-01  
; EARLIER APPLICATION NUMBER: 60/138,629  
; EARLIER FILING DATE: 1999-06-11  
; NUMBER OF SEQ ID NOS: 198  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 96  
; LENGTH: 83  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: SITE  
; LOCATION: (83)  
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids

PCT-US00-15136-96

Query Match 51.4%; Score 423; DB 1; Length 83;  
Best Local Similarity 100.0%; Pred. No. 1.2e-27;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 LPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAIPLPEDAPNAASLAPTPVSPVLEPFNL 134  
|||||  
DB 1 LPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAIPLPEDAPNAASLAPTPVSPVLEPFNL 60  
|||||  
QY 135 TSEPSDYALDLSTFLQOHPAAF 156  
|||||  
DB 61 TSEPSDYALDLSTFLQOHPAAF 82  
|||||

RESULT 11  
PCT-US00-15136-167  
; Sequence 167, Application PC/TUS0015136  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: 50 Human Secreted Proteins  
; FILE REFERENCE: PS551PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/15136  
; CURRENT FILING DATE: 2000-06-01  
; EARLIER APPLICATION NUMBER: 60/138,629  
; EARLIER FILING DATE: 1999-06-11  
; NUMBER OF SEQ ID NOS: 198  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 167  
; LENGTH: 83  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US00-15136-167

Query Match 51.3%; Score 422; DB 1; Length 83;  
Best Local Similarity 98.8%; Pred. No. 1.4e-27;  
Matches 82; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 70 VYRROLPVVEEPNPAKRLFLLLTIVFCQILMAEEGVPAIPLPEDAPNAASLAPTPVSPVL 129  
:|||||  
DB 1 LVYRROLPVVEEPNPAKRLFLLLTIVFCQILMAEEGVPAIPLPEDAPNAASLAPTPVSPVL 60  
|||||  
QY 130 EPNLTSEPSDYALDLSTFLOOH 152  
|||||  
DB 61 EPNLTSEPSDYALDLSTFLOOH 83  
|||||

RESULT 12  
US-60-196-718-4945  
; Sequence 4945, Application US/60196718  
; GENERAL INFORMATION:  
; APPLICANT: Bonazzi, Vivien  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL000456  
; CURRENT APPLICATION NUMBER: US/60/196,718  
; CURRENT FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7494  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4945  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-60-196-718-4945

Query Match 46.3%; Score 381; DB 22; Length 89;  
Best Local Similarity 100.0%; Pred. No. 3.7e-24;  
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRSCHTMTILQAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGRHK 60  
|||||  
DB 1 MCHSRSCHTMTILQAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGRHK 60  
|||||  
QY 61 RSRRLVLYPRV 70  
|||||  
DB 61 RSRRLVLYPRV 70  
|||||

RESULT 13  
PCT-US00-15136-169  
; Sequence 169, Application PC/TUS0015136  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: 50 Human Secreted Proteins  
; FILE REFERENCE: PS551PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/15136  
; CURRENT FILING DATE: 2000-06-01  
; EARLIER APPLICATION NUMBER: 60/138,629  
; EARLIER FILING DATE: 1999-06-11  
; NUMBER OF SEQ ID NOS: 198  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 169  
; LENGTH: 71  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US00-15136-169

Query Match 43.5%; Score 358; DB 1; Length 71;  
Best Local Similarity 97.1%; Pred. No. 2.2e-22;  
Matches 66; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MCHSRSCHTMTILQAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGRHK 60  
|||||  
DB 1 MCHSRSCHTMTILQAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGRHK 60  
|||||  
QY 61 RSRRLVLYP 68  
|||||  
DB 61 RSRRLVLYP 68  
|||||

RESULT 14  
US-60-196-718-5439  
; Sequence 5439, Application US/60196718  
; GENERAL INFORMATION:  
; APPLICANT: Bonazzi, Vivien  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL000456  
; CURRENT APPLICATION NUMBER: US/60/196,718  
; CURRENT FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7494  
; SOFTWARE: FastSeq for Windows Version 4.0



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 19:24:41 ; Search time 36.91 Seconds  
(without alignments)  
75.895 Million cell updates/sec

Title: US-08-799-910-10

Perfect score: 823

Sequence: 1 MCHSRCHPTMTILOAPTPA.....EPDVALDLSTFLQHPAAF 156

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	823	100.0	156	3	US-08-826-246-10
2	823	100.0	156	3	US-08-944-495-10
3	823	100.0	156	3	US-09-126-640-11
4	95.5	11.6	739	3	US-09-035-648-24
5	95.5	11.6	1596	3	US-09-356-952-3
6	93.5	11.4	905	2	US-08-574-959A-9
7	93.5	11.4	1135	2	US-08-574-959A-7
8	92	11.2	1400	1	US-08-080-255-7
9	92	11.2	1400	3	US-08-465-713-7
10	92	11.2	1400	4	PCT-US93-05857-7
11	90.5	11.0	214	1	US-08-217-327-4
12	88	10.7	468	2	US-08-390-000A-7
13	88	10.7	477	1	US-08-444-734A-4
14	88	10.7	477	1	US-08-087-772A-16
15	87	10.6	2414	1	US-08-227-536-2
16	87	10.6	2414	4	PCT-US95-04682-2
17	85.5	10.4	610	1	US-07-821-171B-6
18	85.5	10.4	610	1	US-08-119-262B-6
19	85.5	10.4	610	1	US-08-135-929A-11
20	85.5	10.4	610	1	US-08-234-265A-11
21	85.5	10.4	774	2	US-08-231-193A-42
22	85.5	10.4	774	2	US-08-486-273A-42
23	85.5	10.4	774	3	US-08-480-474-42
24	85.5	10.4	774	3	US-08-940-086A-42
25	85.5	10.4	1214	2	US-08-231-193A-54
26	85.5	10.4	1214	2	US-08-486-273A-54
27	85.5	10.4	1214	3	US-08-480-474-54
28	85.5	10.4	1214	3	US-08-940-086A-54

29	85.5	10.4	1219	2	US-08-231-193A-50
30	85.5	10.4	1219	2	US-08-486-273A-50
31	85.5	10.4	1219	3	US-08-480-474-50
32	85.5	10.4	1219	3	US-08-940-086A-50
33	85.5	10.4	1231	2	US-08-231-193A-48
34	85.5	10.4	1231	2	US-08-486-273A-48
35	85.5	10.4	1231	3	US-08-480-474-48
36	85.5	10.4	1231	3	US-08-940-086A-48
37	85.5	10.4	1236	2	US-08-231-193A-6
38	85.5	10.4	1236	2	US-08-486-273A-6
39	85.5	10.4	1236	3	US-08-480-474-6
40	85.5	10.4	1236	3	US-08-940-086A-6
41	85.5	10.4	1239	2	US-08-231-193A-52
42	85.5	10.4	1239	2	US-08-486-273A-52
43	85.5	10.4	1239	3	US-08-480-474-52
44	85.5	10.4	1239	3	US-08-940-086A-52
45	85.5	10.4	1244	2	US-08-231-193A-46

#### ALIGNMENTS

RESULT 1  
US-08-826-246-10  
; Sequence 10, Application US/08826246  
; Patent No. 6048709  
; GENERAL INFORMATION:  
; APPLICANT: Falb, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/826,246  
; FILING DATE: 28-MAR-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/799,910  
; FILING DATE: 13-FEB-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/011,787  
; FILING DATE: 16-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-078-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)7909090  
; TELEFAX: (212)8699741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 156 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; US-08-826-246-10

Sequence 50, Appl  
Sequence 50, Appl  
Sequence 50, Appl  
Sequence 50, Appl  
Sequence 48, Appl  
Sequence 48, Appl  
Sequence 48, Appl  
Sequence 48, Appl  
Sequence 6, Appl  
Sequence 6, Appl  
Sequence 6, Appl  
Sequence 6, Appl  
Sequence 52, Appl  
Sequence 52, Appl  
Sequence 52, Appl  
Sequence 52, Appl  
Sequence 46, Appl

Query Match 100.0%; Score 823; DB 3; Length 156;  
Best Local Similarity 100.0%; Pred. No. 4e-75;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60  
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60  
Qy 61 RSRRLYPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASL 120  
Db 61 RSRRLYPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASL 120  
Qy 121 APTVPSPVLEPFNLTSPESDYALDLSTFLQOHPAAF 156  
Db 121 APTVPSPVLEPFNLTSPESDYALDLSTFLQOHPAAF 156

## RESULT 2

US-08-944-495-10  
; Sequence 10, Application US/08944495  
; Patent No. 6087477  
; GENERAL INFORMATION:

; APPLICANT: Falb, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/944,495  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/799,910  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:

; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-067-999

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)7909090  
; TELEFAX: (212)8699741

; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 156 amino acids  
; TYPE: amino acid

; STRANDEDNESS: unknown  
; TOPOLOGY: unknown

; MOLECULE TYPE: protein  
; FRAGMENT TYPE: Internal

US-08-944-495-10

Query Match 100.0%; Score 823; DB 3; Length 156;  
Best Local Similarity 100.0%; Pred. No. 4e-75;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60  
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60

Qy 61 RSRRLYPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASL 120  
Db 61 RSRRLYPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASL 120  
Qy 121 APTVPSPVLEPFNLTSPESDYALDLSTFLQOHPAAF 156  
Db 121 APTVPSPVLEPFNLTSPESDYALDLSTFLQOHPAAF 156

## RESULT 3

US-09-126-640-11  
; Sequence 11, Application US/09126640A  
; Patent No. 6099823  
; GENERAL INFORMATION:

; APPLICANT: Falb, Dean A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE  
; FILE REFERENCE: 7853-126  
; CURRENT APPLICATION NUMBER: US/09/126,640A

; EARLIER FILING DATE: 1998-07-30  
; EARLIER APPLICATION NUMBER: 08/870,434  
; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 08/799,910  
; EARLIER FILING DATE: 1997-02-13  
; EARLIER APPLICATION NUMBER: 60/011,787

; EARLIER FILING DATE: 1996-02-16  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11  
; LENGTH: 156  
; TYPE: PRT

; ORGANISM: Homo sapiens  
US-09-126-640-11

Query Match 100.0%; Score 823; DB 3; Length 156;  
Best Local Similarity 100.0%; Pred. No. 4e-75;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60  
Db 1 MCHSRCHPTMTILQAPTPAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK 60

Qy 61 RSRRLYPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASL 120  
Db 61 RSRRLYPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASL 120

Qy 121 APTVPSPVLEPFNLTSPESDYALDLSTFLQOHPAAF 156  
Db 121 APTVPSPVLEPFNLTSPESDYALDLSTFLQOHPAAF 156

## RESULT 4

US-09-035-648-24  
; Sequence 24, Application US/09035648  
; Patent No. 6100031  
; GENERAL INFORMATION:

; APPLICANT: Shyjan, Andrew W.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

; TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL  
; TITLE OF INVENTION: GROWTH AND PROLIFERATION

; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston

; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows95



SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/035,648  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/818,829  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Meiklejohn, Ph.D., Anita L.  
REGISTRATION NUMBER: 35,283  
REFERENCE/DOCKET NUMBER: 07334/003001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 739 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-09-035-648-24

Query Match 11.6%; Score 95.5; DB 3; Length 739;  
Best Local Similarity 27.0%; Pred. No. 0.089; 53; Gaps 7;  
Matches 47; Conservative 14; Mismatches 60; Indels 53; Gaps 7;  
QY 16 APTAPSTIPGP-----RRSGPEITFDPLPEPAAAPAGRPS-----ASRGHR---- 59  
DB 264 APPYSAVTPPDATSGVSSPIAGPAQPPWPQAPWSPAFYDSSERIASRDERISVP 323  
QY 60 -KRS-----RRVLPVRRQLPVEEPNPAKRLFLLL----- 91  
DB 324 AKRTGILQAKRRSTTKPMFTKEPKVSPN--ELLSSLQNSEGKRGTCGAGDGSPEEDY 381  
QY 92 -----TIVFCOILMAEGVAPLPEDAPNAASLAP-TPVSPVLEPFLNLTSEP 138  
DB 382 LSLGAEACNFMOSSAKQKTPPPVAPKPAVKSSSQPVTPVSPVWSPGVAPTOP 435

RESULT 5  
US-09-356-952-3  
Sequence 3, Application US/09356952  
Patent No. 6117663  
GENERAL INFORMATION:  
APPLICANT: Boriack-Sjodin, Ann  
APPLICANT: Margarit, S. M.  
APPLICANT: Bor-Sogli, Dafna  
APPLICANT: Cole, Philip  
APPLICANT: Kurlyan, John  
TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE  
FILE REFERENCE: 600-1-228N  
CURRENT APPLICATION NUMBER: US/09/356,952  
EARLIER FILING DATE: 1999-07-19  
EARLIER APPLICATION NUMBER: 60/093,631  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 1596  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-09-356-952-3

Query Match 11.6%; Score 95.5; DB 3; Length 1596;  
Best Local Similarity 28.3%; Pred. No. 0.22;  
Matches 52; Conservative 19; Mismatches 68; Indels 45; Gaps 13;

QY 4 SRSCHPTWTLQATPPAPSTIPGHRG--SCPEITFDPLPEPAAAPAGRPSASKGHRK 61  
DB 1354 TESCADMAQKRAAP-DAP-TLP-PRDGELSP-----PIPPRLNIHSTGISYLRSHGKS 1404  
QY 62 SRRV-----LYPR---VVRQLPVE-----BPNPAKRLFLLLTIVFCOILMAEGV 105  
DB 1405 KEFVGNSSLLLPNTSSIMIRNSALEKRAAATSPNQAAAGPISTTTLVTVSOAVATDEPL 1464  
QY 106 PAPLPEDAPNAASLAPT--PVSPVLEPF--NLTSEP-----SYALDLSTFLQO----H 152  
DB 1465 PLPISP-----AASSSTTTSPLTPAMSPMSPNIPSHPVVESTSSSVYHQLRMKQOQOQTH 1519  
QY 153 PAAP 156  
DB 1520 PAIV 1523

RESULT 6  
US-08-574-959A-9  
Sequence 9, Application US/08574959A  
Patent No. 5962224  
GENERAL INFORMATION:  
APPLICANT: Jaekyoon Shin, Insil Joong, Rauna K. Vadiamudi  
APPLICANT: and Jack L. Strominger  
TITLE OF INVENTION: P62 POLYPEPTIDES, RELATED POLYPEPTIDES  
TITLE OF INVENTION: AND USES THEREFOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, Suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/574,959A  
FILING DATE: 19-DEC-95  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: DFN-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 905 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-574-959A-9

Query Match 11.4%; Score 93.5; DB 2; Length 905;  
Best Local Similarity 23.1%; Pred. No. 0.18;  
Matches 45; Conservative 17; Mismatches 62; Indels 71; Gaps 8;  
QY 8 HPTMTILQ-----APTAPSTIPGPRGSGPE-----LFTFDPLPEPAAAPAGRPSAS 55  
DB 400 HFRVPLQPMGPTCTPAPVPLLRPHRPSGPHRSILRAPCPQWAPCPQAPCPGAPGMS 459  
QY 56 RG-----HKKRSRR--VLYP----- 68  
DB 460 AGPVSEPWTSTTANLLGLLSRPSVCPPLLPENHNRAGSNEDPIAPSGTPPTIPD 519  
QY 69 -----RVVRQLPVEEPNPAKRLFLLLTIVFCOILMAEGVAPLPEDAPNAASLAPT 123  
DB 520 ETFGGRVPRPAFVHVHDKESADSVETISLSDSDSVVIVPEGLP-PLPP---PPPSGATPP 575



```
;
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465.713
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/080.255
; FILING DATE: 17 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US93-05857-7
;
; Query Match 11.2%; Score 92; DB 3; Length 1400;
; Best Local Similarity 22.1%; Pred. No. 0.42;
; Matches 31; Conservative 17; Mismatches 40; Indels 52; Gaps 5;
;
QY 15 QAPTPAPSTIPGPRGSGPEITFDPLPE-----PAAAPAGRPSASRGHRKKS 62
DB 465 QKPTPSAREDPAPKKSSS-----EPPPRKPVEEKSEGCNVSAPGCSKQATTPASRKSS 518
QY 63 RVLVPRVV-----RRQLPVEEPNPAKRLLFLLLTVFCOILMAEEGVAPAPLPP 111
DB 519 KOVSOPALVIPPQPTTGPRKEVKTTPSEPKK-----KQPPPP 558
QY 112 EDAPNAA---SLAPTPVSPV 128
DB 559 ESGPEQSKQKKVAPRPSIPV 578
;
RESULT 10
PCT-US93-05857-7
; Sequence 7, Application PC/TUS9305857
; GENERAL INFORMATION:
; APPLICANT: Board of Regents
; APPLICANT: The University of Texas System
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: GENE REARRANGEMENTS AND TRANSLOCATIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05857
; FILING DATE: 19930617
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
```

```
;
; APPLICATION NUMBER: 07/900.689
; FILING DATE: 17/06/92
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US93-05857-7
;
; Query Match 11.2%; Score 92; DB 4; Length 1400;
; Best Local Similarity 22.1%; Pred. No. 0.42;
; Matches 31; Conservative 17; Mismatches 40; Indels 52; Gaps 5;
;
QY 15 QAPTPAPSTIPGPRGSGPEITFDPLPE-----PAAAPAGRPSASRGHRKKS 62
DB 465 QKPTPSAREDPAPKKSSS-----EPPPRKPVEEKSEGCNVSAPGCSKQATTPASRKSS 518
QY 63 RVLVPRVV-----RRQLPVEEPNPAKRLLFLLLTVFCOILMAEEGVAPAPLPP 111
DB 519 KOVSOPALVIPPQPTTGPRKEVKTTPSEPKK-----KQPPPP 558
QY 112 EDAPNAA---SLAPTPVSPV 128
DB 559 ESGPEQSKQKKVAPRPSIPV 578
;
RESULT 11
US-08-217-327-4
; Sequence 4, Application US/08217327
; Patent No. 5474925
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E
; APPLICANT: Barton, Kenneth A
; TITLE OF INVENTION: Immobilized proteins in Cotton Fiber
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: P.O. Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/217,327
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/812,233
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 214 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-217-327-4

Query Match 11.0% Score 90.5; DB 1; Length 214;  
Best Local Similarity 27.5% Pred. No. 0.066;

Matches 38; Conservative 8; Mismatches 59; Indels 33; Gaps 5;

QY 9 PTMTILOQTPAPSTIPGRRGSGEIFTDP-----LPEAAAPAGRPSASRCHRRKS 62

DB 54 PTATPPPVSTPTPTSSPPVPTASPPVSTPPSPPPATPPASPPATPPASPPATP 113

QY 63 RRVLYPRVRLQPLVEEPNPAKRLLELLLTIVFCQILMAEGVPAPLPPEADAPNAASLAP 122

DB 114 PPASPPATP---PPATPPA-----TPPATPPAPPLASPPAT 149

QY 123 TP-VSPVLEPNTSEPS 139

DB 150 VPASPVQTP--LTSPPA 165

RESULT 12

US-08-390-000A-7

; Sequence 7, Application US/08390000A

; Patent No. 5985363

; GENERAL INFORMATION:

; APPLICANT: Sealion, Stuart C.

; TITLE OF INVENTION: Cloning and Expression of

; TITLE OF INVENTION: Gonadotropin-Releasing Hormone Receptor

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds LLP

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/390,000A

; FILING DATE: 17-FEB-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Mirock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 6923-052

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 790-9090

; TELEFAX: 212 869-8864/9741

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 468 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

US-08-390-000A-7

Query Match 10.7% Score 88; DB 2; Length 468;

Best Local Similarity 24.2% Pred. No. 0.29;

Matches 50; Conservative 19; Mismatches 44; Indels 94; Gaps 12;

QY 17 PTPAPSTIPGRRGSGEIFTDPPLPEPAAAPAGRPSAS--RGHKRSRRVLYPRVRRQ 74

DB 264 PPSPSPPVAPAPPVGP-----PRPAAAAATAPLANGRAGRPPSRSLV----ALREQ 311

QY 75 -----LPVEEPNPAKH-----LPVCO- 97

DB 312 KAKTLGIIMGVFTLCWLPPFLANVKAHRELVPDLFVFNLMGYANSFNIYCRS 371

QY 98 --ILMAEGV-----PAPLPPEADAPNAAS-----LA 121

DB 372 PDFRKAQGLCCARRARRRHATHGDRPRASGCLARPGP-PP-SYCAASDUUUUUUVG 428

QY 122 PTPVSPVLEPF-----NLTSEPSDYALD 144

DB 429 ATPPARLLEPWAGCGGAAADSDSLD 455

RESULT 13

US-08-444-734A-4

; Sequence 4, Application US/08444734A

; Patent No. 5610282

; GENERAL INFORMATION:

; APPLICANT: Sibley, David R.

; APPLICANT: Monsma, Frederick J.

; APPLICANT: Mahan, Lawrence C.

; APPLICANT: McVittie, Loris D.

; TITLE OF INVENTION: cDNA encoding the rat D1 dopamine

; TITLE OF INVENTION: receptor linked to adenylyl cyclase activation and

; TITLE OF INVENTION: expression of the receptor protein in plasmid-transfect

; TITLE OF INVENTION: cell lines

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe, Martens, Olson and Bear

; STREET: 620 Newport Center Drive, Sixteenth Floor

; CITY: Newport Beach

; STATE: CA

; COUNTRY: USA

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/444,734A

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/029,917

; FILING DATE: 03-MAR-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/548,714

; FILING DATE: 06-JUL-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Altman, Daniel E.

; REGISTRATION NUMBER: 34,115

; REFERENCE/DOCKET NUMBER: NIH065.001FWJ

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (714) 760-0404

; TELEFAX: (714) 760-9502

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 477 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: internal

US-08-444-734A-4

Query Match

Best Local Similarity

Matches

10.7% Score 88; DB 1; Length 477;

24.2% Pred. No. 0.3;

50; Conservative 19; Mismatches 44; Indels 94; Gaps 12;

[illegible]

Search completed: January 30, 2001, 22:06:25  
Job time: 9704 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 18:47:31 ; Search time 54.97 Seconds  
(without alignments)  
97.039 Million cell updates/sec

Title: US-08-799-910-10  
Perfect score: 823  
Sequence: 1 MCHSRCHPTMTILQAPTA.....EPSDYALDLSTFLQHPAAF 156

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_36:\*\*  
1: /cgn2\_2/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /cgn2\_2/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /cgn2\_2/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /cgn2\_2/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /cgn2\_2/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /cgn2\_2/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /cgn2\_2/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /cgn2\_2/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /cgn2\_2/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /cgn2\_2/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /cgn2\_2/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /cgn2\_2/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /cgn2\_2/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /cgn2\_2/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /cgn2\_2/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /cgn2\_2/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /cgn2\_2/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /cgn2\_2/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /cgn2\_2/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /cgn2\_2/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /cgn2\_2/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	823	100.0	156	21	Y45017	Protein encoded by
2	816	99.1	156	18	W36006	Human fchd605 gene
3	107	13.0	206	18	W14574	Streptococcus pneu
4	104	12.6	129	18	W14573	Streptococcus pneu
5	101.5	12.3	190	18	W14569	Streptococcus pneu
6	101	12.3	180	18	W14562	Streptococcus pneu
7	101	12.3	183	18	W14570	Streptococcus pneu
8	100	12.2	783	19	W37151	Streptococcus pneu
9	100	12.2	787	19	W37152	Mouse neural Mena+
10	100	12.2	802	19	W37153	Mouse neural Mena+
11	98.5	12.0	168	21	Y68732	Amino acid sequenc
12	96.5	11.7	179	21	Y68731	Amino acid sequenc

13	95.5	11.6	739	18	W25790	Gene 036 product d
14	95.5	11.6	1596	21	Y68821	Amino acid sequenc
15	94.5	11.5	520	16	R78185	Protein sequence o
16	94	11.4	1290	20	Y06809	Human Per gene pro
17	94	11.4	1290	21	Y32216	Human PER protein.
18	94	11.4	1291	20	Y01687	Protein encoded by
19	93.5	11.4	185	18	W14566	Streptococcus pneu
20	93.5	11.4	905	18	W31186	Human p160 polypep
21	93.5	11.4	1135	18	W31185	Human p160 polypep
22	92.5	11.2	572	18	W31855	Mycobacterium tube
23	92.5	11.2	763	18	W31852	Mycobacterium tube
24	92	11.2	166	18	W14558	Mycobacterium tube
25	92	11.2	167	18	W14558	Streptococcus pneu
26	92	11.2	1400	15	W44514	Streptococcus pneu
27	92	11.2	1400	17	R92705	MLL amino acid seq
28	92	11.2	3910	14	R38470	MLL cDNA clone 14-
29	92	11.2	3910	16	R66462	ALL-1 protein. Ho
30	91	11.1	1012	20	Y17406	Human atrophin-1 r
31	90.5	11.0	214	17	R86913	Cotton fiber-speci
32	90	10.9	171	20	Z29034	T. gondii immunoge
33	90	10.9	171	20	Y29074	T. gondii immunoge
34	90	10.9	254	20	Z29061	T. gondii immunoge
35	89.5	10.9	666	19	W72911	Mycobacterium tube
36	89.5	10.9	666	20	Y21928	Amino acid sequenc
37	89	10.8	788	21	Y54466	Amino acid sequenc
38	89	10.8	807	21	Y54467	Protein encoded by
39	89	10.8	1291	20	Y01689	Mouse Per gene pro
40	89	10.8	1291	20	Y06810	Product of the cDN
41	89	10.8	3969	15	R52971	Streptococcus pneu
42	88.5	10.8	187	18	W14579	Streptococcus pneu
43	88.5	10.8	188	18	W14580	Human transmembran
44	88.5	10.8	688	21	Y57898	Human transcriptio
45	88.5	10.8	711	19	W80602	

ALIGNMENTS

RESULT 1  
Y45017  
ID Y45017 standard; Protein; 156 AA.  
AC Y45017;  
XX  
XX  
DT 31-MAY-2000 (first entry)  
XX  
DE Protein encoded by fchd605 gene.

fchd605 gene; human; cardiovascular disease; oncogenic disorder;  
diabetic retinopathy; fibroproliferative disorder; atherosclerosis;  
TGF-beta signalling pathway; TGF; Transforming growth factor;  
pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;  
vascularisation; cytostatic; antidiabetic; ophthalmological.

OS Homo sapiens.  
XX  
XX  
PN WO200006206-A1.  
XX  
PD 10-FEB-2000.  
XX  
PF 30-JUL-1999; 99WO-US17394.  
XX  
PR 30-JUL-1998; 98US-0126640.  
XX  
PA (MILL-) MILLENNium PHARM INC.  
XX  
PI Falb DA;  
XX  
DR WPI: 2000-205414/18.  
XX  
DR N-PSDB: 250711.

PT Identifying substances for ameliorating symptoms of fibroproliferative diseases or oncogenic related disorders -

XX

PS Example: Fig 5; 214pp; English.

XX

CC The patent discloses methods for the treatment and diagnosis of cardiovascular diseases by novel human genes which are differentially expressed in different cardiovascular disease states. Compositions which can modify MCP-beta signalling pathway are identified by screening. CC These are used therapeutically to treat fibroproliferative and oncogenic disorders, especially TGF (transforming growth factor)-beta related disorders, including diabetic retinopathy, arteriosclerosis, pancreatic cancer, angiogenesis, inflammation, fibrosis, tumour growth and vascularisation. The present sequence is the protein product of fchd605 gene which is up-regulated in monocytes treated with oxidised LDL (low density lipoprotein) can be used to design cardiovascular disease treatment strategies. Depending on whether the up-regulation has a pathogenic or protective effect treatment methods can be designed to increase or decrease the activity of the protein product of the gene.

XX

SQ Sequence 156 AA;

Query Match 100.0%; Score 823; DB 21; Length 156;  
Best Local Similarity 100.0%; Pred. No. 7.4e-69;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEITFDPLPEPAAAPACRPSASRGHRK 60  
|||||  
DB 1 mchsrchpmtlilqaptapstipgprgsgpeitfdplpepaaapagrpasrghrk 60  
|||||  
QY 61 RSRRLVYPRVVRQIPVEEPNPAKRLLFLLITIVFCQILMAEGVPAPILPPEDAPNAASL 120  
|||||  
DB 61 rsrriyprvrrqipveepnpakrllfllitvfcqilmaegvpapilppedapnaasl 120  
|||||  
QY 121 APTVPSPVLEPFNLTSEPSDYALDLSTFLQHPAAF 156  
|||||  
DB 121 aptvpvlepfntsepsdyaldlstflqhpaaaf 156  
|||||

RESULT 2

W36006

ID W36006 standard; Protein; 156 AA.

XX W36006;

DT 03-MAR-1998 (first entry)

DE Human Fchd605 gene product.

XX Fchd605 gene; differential expression; monocyte; human;  
KW foam cell; cardiovascular disease; arteriosclerosis; ischaemia;  
KW reperfusion; hypertension; restenosis; arterial inflammation;  
KW therapy; diagnosis; drug screening; marker.

XX Homo sapiens.

XX WO9730065-A1.

XX 21-AUG-1997.

XX 14-FEB-1997; 97WO-US02291.

XX 13-FEB-1997; 97US-0799910.

XX 16-FEB-1996; 96US-0011787.

XX (MILL-) MILLENNIUM PHARM INC.

XX Falb DA;

XX WPI: 1997-424966/39.

XX N-PSDB: T94471.

XX New genes differentially expressed in cardiovascular disease - used  
PT for diagnosis, drug screening and treatment of cardiovascular

PT disease, e.g. atherosclerosis, restenosis, hypertension, etc

XX Example 6; Fig 5; 163pp; English.

XX This protein is encoded by the novel human fchd605 gene (see T94471) that is up-regulated in monocytes treated with oxidised low density lipoproteins that simulate the conditions under which foam cells develop during atherogenesis. The protein has sequence similarity to the mouse gly96 gene and to EST T49532. Novel fchd531, fchd540, fchd545, fchd602 and fchd605 genes (see T94467-71) provide a fingerprint for the study of cardiovascular diseases, including atherosclerosis, ischaemia/reperfusion, hypertension, restenosis and arterial inflammation. Methods are provided for the diagnosis, monitoring in clinical trials, screening for therapeutically effective compounds, and treatment of cardiovascular diseases based on discoveries regarding the expression patterns of these novel genes.

XX Sequence 156 AA;

Query Match 99.1%; Score 816; DB 18; Length 156;  
Best Local Similarity 99.4%; Pred. No. 3.3e-68;  
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILQAPTAPSTIPGPRGSGPEITFDPLPEPAAAPACRPSASRGHRK 60  
|||||  
DB 1 mchsrchpmtlilqaptapstipgprgsgpeitfdplpepaaapagrpasrghrk 60  
|||||  
QY 61 RSRRLVYPRVVRQIPVEEPNPAKRLLFLLITIVFCQILMAEGVPAPILPPEDAPNAASL 120  
|||||  
DB 61 rsrriyprvrrqipveepnpakrllfllitvfcqilmaegvpapilppedapnaasl 120  
|||||  
QY 121 APTVPSPVLEPFNLTSEPSDYALDLSTFLQHPAAF 156  
|||||  
DB 121 aptvpvlepfntsepsdyaldlstflqhpaaaf 156  
|||||

RESULT 3

W14574

ID W14574 standard; Protein; 206 AA.

XX W14574;

DT 28-OCT-1997 (first entry)

DE Streptococcus pneumoniae PspA central region.

XX PspA; pneumococcal surface protein; vaccine; otitis media;  
KW meningitis; bacteraemia; pneumonia.  
XX Streptococcus pneumoniae strain Dbl5.

XX Key Location/Qualifiers

XX Misc-difference 50 /note= "unidentified amino acid"

XX WO9709994-A1.

XX 20-MAR-1997.

XX 16-SEP-1996; 96WO-US14819.

XX 15-SEP-1995; 95US-0529055.

XX (UABR-) UAB RES FOUND.

XX Briles DE, Brooks-Walter A, Crain MJ, Hollingshead S;  
PI McDaniel LS, Swiatlo E, Tart R, Yother J;  
XX WPI: 1997-202002/18.

XX Streptococcus pneumoniae surface protein PspC and truncated PspA -  
PT







XX  
PT - used in control of cytoskeletal dynamic events in normal and  
pr abnormal cell morphology, adhesion, motility, growth and  
pr differentiation  
XX  
PS Example 4; Page 58-60; 77pp; English.  
XX  
CC This protein comprises novel murine neural Menat. Its amino acid  
CC sequence was deduced from a cDNA clone (see V02998) obtained from

Gertler FB, Niebuhr K, Soriano P, Wehland J;  
 WPI: 1998-101197/09.  
 Detection of modulators of Mena and Ena-VASP-like genes and proteins  
 - used in control of cytoskeletal dynamic events in normal and  
 abnormal cell morphology, adhesion, motility, growth and  
 differentiation  
 Example 4: Page 60-63; 77pp: English.

CC This protein comprises novel murine neural Mena+. Its amino acid  
 CC sequence was deduced from a cDNA clone obtained from a mouse brain  
 CC cDNA library. Two other isoforms, neural Mena+ (see W37151) and  
 CC neural Mena++ (see W37153), are also disclosed. Unlike mammalian  
 CC Ena (Mena, see W37148), neural Mena isoforms exhibit neural  
 CC tissue-specific distribution. Based on the disclosed Mena and Evi  
 CC genes (see also V02996-98) and proteins (see also W37148-49), a  
 CC variety of methods and compositions are provided for screening,  
 CC isolating and characterising endogenous and exogenous factors,  
 CC drugs and therapeutic agents useful to evaluate and/or control  
 CC cytoskeletal dynamic events involved in normal and abnormal cell  
 CC morphology, adhesion, motility, growth and/or differentiation. A  
 CC method of detecting a modulator of Mena activity/expression is  
 CC claimed.

XX Sequence 787 AA;

Query Match 12.2%; Score 100; DB 19; Length 787;

Best Local Similarity 26.0%; Pred. No. 0.21; Mismatches 51; Indels 46; Gaps 6;

Matches 38; Conservative 11;

Oy 2 CHRSCHPTMT-ILQAPTPAPSTIPGPRGSGPEIFDPLPFAAAP-----ACRPSASR 56

Db 456 cgsqaspppgtplastpskpsvlpssaga-----pasaetplnplgdssase 505

Oy 57 GHRKRRVLYPRVVRQLPVEEPNPAKRLFLILLITIVFCQLMAECSVPAPLPEDAPN 116

Db 506 -----pglqaasqpaesptpglglv-----gppappppplps 538

Oy 117 ----AASLAPTPVSPVLEPNLTSEP 138

Db 539 gpayasalpppppppppppplstgpp 564

RESULT 10

ID W37153 standard; Protein; 802 AA.

AC W37153;

DT 06-JUL-1998 (first entry)

DE Mouse neural Mena+++ protein.

KW Neural Mena+++ protein; mammalian Ena; Enabled protein; Evi protein;

KW Cytoskeleton; cell morphology; cell adhesion; cell differentiation;

KW cell growth; cell motility; mouse.

OS Mus musculus.

PN W09801755-A1.

PD 15-JAN-1998.

PF 03-JUL-1997; 97WO-US11669.

PR 05-JUL-1996; 96US-0675815.

PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.

PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.

XX Gertler FB, Niebuhr K, Soriano P, Wehland J;

XX WPI; 1998-101197/09.

DR Detection of modulators of Mena and Ena-VASP-like genes and proteins

PT - used in control of cytoskeletal dynamic events in normal and

PT abnormal cell morphology, adhesion, motility, growth and

PT differentiation

XX Example 4; Page 63-65; 77pp; English.

XX

CC This protein comprises novel murine neural Mena+. Its amino acid  
 CC sequence was deduced from a cDNA clone obtained from a mouse brain  
 CC cDNA library. Two other isoforms, neural Mena+ (see W37151) and  
 CC neural Mena++ (see W37152), are also disclosed. Unlike mammalian  
 CC Ena (Mena, see W37148), neural Mena isoforms exhibit neural  
 CC tissue-specific distribution. Based on the disclosed Mena and Evi  
 CC genes (see also V02996-98) and proteins (see also W37148-49), a  
 CC variety of methods and compositions are provided for screening,  
 CC isolating and characterising endogenous and exogenous factors,  
 CC drugs and therapeutic agents useful to evaluate and/or control  
 CC cytoskeletal dynamic events involved in normal and abnormal cell  
 CC morphology, adhesion, motility, growth and/or differentiation. A  
 CC method of detecting a modulator of Mena activity/expression is  
 CC claimed.

XX Sequence 802 AA;

Query Match 12.2%; Score 100; DB 19; Length 802;

Best Local Similarity 26.0%; Pred. No. 0.21;

Matches 38; Conservative 11; Mismatches 51; Indels 46; Gaps 6;

Oy 2 CHRSCHPTMT-ILQAPTPAPSTIPGPRGSGPEIFDPLPFAAAP-----ACRPSASR 56

Db 471 cgsqaspppgtplastpskpsvlpssaga-----pasaetplnplgdssase 520

Oy 57 GHRKRRVLYPRVVRQLPVEEPNPAKRLFLILLITIVFCQLMAECSVPAPLPEDAPN 116

Db 521 -----pglqaasqpaesptpglglv-----gppappppplps 553

Oy 117 ----AASLAPTPVSPVLEPNLTSEP 138

Db 554 gpayasalpppppppppppplstgpp 579

RESULT 11

Y68732

ID Y68732 standard; Protein; 168 AA.

AC Y68732;

DT 05-MAY-2000 (first entry)

DE Amino acid sequence of high mobility group (HMG)-Y/R protein.

KW Retinoic acid receptor; RAR; high mobility group protein;

KW HMG protein; nuclear receptor; antiproliferative; anti-inflammatory;

KW antipsoiasis; anticancer.

OS Homo sapiens.

PN W0200001717-A2.

PD 13-JAN-2000.

PF 28-JUN-1999; 99WO-US14801.

PR 01-JUL-1998; 98US-0108298.

PA (ALLR) ALLERGAN SALES INC.

PA Nagpal S, Chandraratna RA, Ghosh C;

XX WPI; 2000-170999/15.

DR Identifying modulators of nuclear receptors, useful e.g. as anticancer

PT and anti-inflammatory agents, according to their effect on binding

PT between the receptor and high-mobility group protein

XX Claim 28; Fig 1B; 43pp; English.

XX The present sequence represents a high mobility group (HMG)-Y/R protein.

CC HMG sequences that interact with retinoic acid receptors (RARs) in a

CC Ligand dependent manner to induce gene expression can be identified  
 CC using the method of the invention. The specification describes a  
 CC method for identifying compounds that modulate a nuclear receptor.  
 CC The method comprises incubating a polypeptide containing a  
 CC ligand-binding region of nuclear receptor with test compound in  
 CC presence of second polypeptide containing the nuclear receptor-binding  
 CC part of a HMG protein, and detecting any alteration in binding between  
 CC the two polypeptides. The method is useful for identifying potential  
 CC antiproliferative, anti-inflammatory, antipsoriasis and anticancer  
 CC agents.

XX  
 SQ Sequence 168 AA;

Query Match 12.0%; Score 98.5; DB 21; Length 168;

Best Local Similarity 27.7%; Pred. No. 0.05;

Matches 39; Conservative 9; Mismatches 52; Indels 41; Gaps 4;

OY 17 PTPAPSTIPGPRGSGPEIFTDPLPEPAAPACRPASAKGHRKSRRLVPRVRRQL- 75

DB 33 ppkpeveptpkr-----prgpkyskknwrrkrkrasrrprrrss 74

OY 76 -PVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASIAPTVSPVLEPENL 134

DB 75 dpcvpapahwrssflilgdsf-----aplp-----pppqpahhhrl 113

OY 135 PSEPSDYALDLSFTFLQHPAA 155

DB 114 wpppsstcaltttlhstpaa 134

RESULT 12

ID Y68731

AC Y68731 standard; Protein; 179 AA.

XX

XX

XX

DT 05-MAY-2000 (first entry)

DE Amino acid sequence of high mobility group (HMG)-R protein.

XX Retinoic acid receptor; RAR; high mobility group protein;

KW HMG protein; nuclear receptor; antiproliferative; anti-inflammatory;

KW antipsoriasis; anticancer.

XX

XX

OS Homo sapiens.

XX WO200001717-A2.

PN 13-JAN-2000.

PD 28-JUN-1999; 99WO-US14801.

XX 01-JUL-1998; 98US-0108298.

XX (ALLR ) ALLERGAN SALES INC.

XX Nagpal S, Chandratatna RA, Ghosh C;

XX WPI: 2000-170999/15.

DR Identifying modulators of nuclear receptors, useful e.g. as anticancer

PT and anti-inflammatory agents, according to their effect on binding

PT between the receptor and high-mobility group protein

XX Claim 24; Fig 1B; 43pp; English.

XX The present sequence represents a high mobility group (HMG)-R protein.

CC HMG sequences that interact with retinoic acid receptors (RARs) in a

CC ligand dependent manner to induce gene expression can be identified

CC using the method of the invention. The specification describes a

CC method for identifying compounds that modulate a nuclear receptor.

CC The method comprises incubating a polypeptide containing a

CC ligand-binding region of nuclear receptor with test compound in  
 CC presence of second polypeptide containing the nuclear receptor-binding  
 CC part of a HMG protein, and detecting any alteration in binding between  
 CC the two polypeptides. The method is useful for identifying potential  
 CC antiproliferative, anti-inflammatory, antipsoriasis and anticancer  
 CC agents.

XX  
 SQ Sequence 179 AA;

Query Match 11.7%; Score 96.5; DB 21; Length 179;

Best Local Similarity 27.4%; Pred. No. 0.083;

Matches 40; Conservative 9; Mismatches 56; Indels 41; Gaps 4;

OY 12 TILQAPTAPSTIPGPRGSGPEIFTDPLPEPAAPACRPASAKGHRKSRRLVPRVV 71

DB 39 talvsgqkpeveptpkr-----prgpkyskknwrrkrkrasrrsp 80

OY 72 RRQL--PVEEPNPAKRLFLLLTIVFCQILMAEGVPAPLPEDAPNAASLAPTVSPVL 129

DB 81 rrrssdpcvpapahwrssflilgdsf-----aplp-----pppqpah 119

OY 130 EFPNLTSEPSDYALDLSFTFLQHPAA 155

DB 120 hhrllwpppsstcaltttlhstpaa 145

RESULT 13

W25790

ID W25790 standard; Protein; 739 AA.

XX

XX

XX

DT 11-MAR-1998 (first entry)

XX Gene 036 product differentially expressed in colon tumour cells.

XX Colon tumour; colon cancer; differential expression; gene 036;

XX human; diagnosis; gene therapy; tumour suppressor.

XX Homo sapiens.

XX WO9733551-A2.

XX 18-SEP-1997.

XX 14-MAR-1997; 97WO-US04191.

XX 15-MAR-1996; 96US-0013438.

XX (MILL-) MILLENNIUM PHARM.

XX Shyjan AW;

XX WPI: 1997-470615/43.

XX N-PSDB; T91708.

XX Gene 036 with reduced or amplified expression in tumour cells -

XX used, optionally with genes 097, 030 and 036 or their protein

XX products, for diagnosis and treatment of colonic cancer

XX Claim 38; Fig 2; 117pp; English.

XX This protein is encoded by gene 036 (see T91708), which was

XX identified by differential display analysis as being expressed at

XX a higher level in normal colon tissues than in cancerous colon

XX tissues. Gene 036 is a candidate tumour suppressor gene. A

XX correlation was found between an increase in the expression level

XX of gene 036 and a decrease in a colon cell's tumour potential

XX Hence, methods that increase the level of expression of gene 036

XX may inhibit or slow the progression to tumours and cancer, e.g.

XX colon cancer. The 036 protein is used in claimed methods for

XX treating a patient suffering from a disorder associated with

CC insufficient expression of gene 036 protein, and for identifying  
CC compounds that modulate 036 protein activity. Such compounds are  
CC useful in the diagnosis, prevention and treatment of tumours and  
CC cancers.

XX Sequence 739 AA;

Query Match 11.6%; Score 95.5; DB 18; Length 739;  
Best Local Similarity 27.0%; Pred. NO. 0.5;  
Matches 47; Conservative 14; Mismatches 60; Indels 53; Gaps 7;  
QY 16 APTAPSTIPGP-----RRGSGPEIFTFPLPEPAAPAGRPDS-----ASRGHR----- 59  
DB 264 appysavtppdpafgrvssplagpaggppwpqapwsqpfydsseriesdrisvp 323  
QY 60 -RRS-----RRVLYPRVRRQLPVEEFPNPAKRLFLLL----- 91  
DB 324 akrtgilqearstutkpmftfkpkvsnpp--ellsllqnsegkrgtgaggdspeedy 381  
QY 92 -----TIVFCQIILMAEGVPAPLPPEDAPNAASIAP-TPVSPVLEPNLTSEP 138  
DB 382 lsiqaacnfmqssakqtkppvpakpvaksssgqvtptvpsvpsvqvaptqp 435

RESULT 14

Y68821  
ID Y68821 standard; protein; 1596 AA.  
XX Y68821;  
AC Y68821;  
DT 16-MAY-2000 (first entry)  
XX  
DE Amino acid sequence of a Drosophila Son of sevenless (Sos) protein.  
XX  
KW Ras; Son of sevenless; Sos; crystal; Ras-Sos complex; cancer;  
KW protein coordinate data.  
XX  
OS Drosophila melanogaster.  
XX  
PN W0200005258-A1.  
XX  
PD 03-FEB-2000.  
XX  
PF 20-JUL-1999; 99WO-US16348.  
XX  
PR 21-JUL-1998; 98US-0119794.  
XX  
PA (UYRQ ) UNIV ROCKEFELLER.  
XX  
PI Horiack-Sjodin A, Margarit SM, Bar-Sagi D, Cole P, Kuriyan J;  
XX  
DR WPI; 2000-182647/16.  
XX  
PT Novel crystals comprising a Ras-Son of sevenless complex, useful for  
PT screening drugs useful in cancer treatment  
XX  
PS Disclosure: Page 169-175; 224pp; English.

XX The specification describes a crystal complex comprising at least a  
CC Sos contacting region of a Ras protein and at least a Son of sevenless  
CC (Sos) protein catalytic region fragment, that effectively diffracts  
CC X-rays. Ras and Sos form a tight complex. Sos does not impede the  
CC binding sites for the nucleotide base and the ribose of GTP or GDP  
CC and thus the Ras-Sos complex maintains a structure that permits  
CC nucleotide release and rebinding. The crystals are used for the  
CC determination of the atomic coordinates of the complex to a resolution  
CC of more than 5.0 Angstrom. The crystals, or a dataset comprising the  
CC three-dimensional coordinates obtained from the crystals, is useful  
CC for identifying an agent that stabilizes the Ras-Sos complex. The  
CC crystals are also useful for identifying agents that inhibit the  
CC formation of Ras-Sos complex. Ras and Sos fragments are useful for  
CC growing a crystal of a protein-ligand complex. Agents that stabilize

CC or inhibit the formation of Ras-Sos complex are useful in the  
CC treatment of cancer. The present sequence represents a Drosophila  
CC Sos protein.

XX Sequence 1596 AA;

Query Match 11.6%; Score 95.5; DB 21; Length 1596;  
Best Local Similarity 28.3%; Pred. No. 1.2;  
Matches 52; Conservative 19; Mismatches 68; Indels 45; Gaps 13;  
QY 4 SRSCHPTMTILQAPTAPSTIPGPRRG--SGPEIFTFPLPEPAAPAGRPASRGRHKR 61  
DB 1354 tescadmqrqap-dap-clp-prdgelspp-----pippclnhstgisylrqshgks 1404  
QY 62 SRRV-----LYPR-----VVRQLPVE-----EPNPAKRLFLLLTIVFCQIILMAEGV 105  
DB 1405 kefvgnssllpntssimicrnaslekrasatqpnqaaagpisttvtvsqvatdepl 1464  
QY 106 PAPLPEDAPNAASLAPT--PVSPVLEPF--NLTSEP-----SDYALDLSTFLQO----H 152  
DB 1465 plpisp-----aassttsltptampspnpsphvestsssyahqlrmrqgqqqth 1519  
QY 153 PAAP 156  
DB 1520 paip 1523

RESULT 15

R78185  
ID R78185 standard; Protein; 520 AA.  
XX R78185;  
AC R78185;  
DT 09-FEB-1996 (first entry)  
XX  
DE Protein sequence of PEA3-beta -an ETS transcription factor.  
XX  
KW Transcription factor; probe; reverse transcription; PCR; primer;  
KW expression vector; E.coli; COS cell; ras; cancer cell multiplication;  
KW polyoma virus; transformation.  
XX  
OS Homo sapiens.  
XX  
PH Key Location/Qualifiers  
FT Misc-difference 84 /note= "encoded by CTC"  
FT Misc-difference 126 /note= "encoded by AAC"  
FT Misc-difference 500 /note= "encoded by ACC"  
FT  
XX JP07145197-A.  
XX  
XX 06-JUN-1995.  
XX  
XX 25-NOV-1993; 93JP-0295393.  
XX  
XX 25-NOV-1993; 93JP-0295393.  
XX  
XX (EISA ) EISAI CO LTD.  
XX (HIRA/) HIRANO T.  
XX  
XX WPI; 1995-237197/31.  
XX N-P5DB; Q91769.  
XX  
PT ETS transcription factor activated by ras - may be used in the study  
PT of cancer cell proliferation and the proliferation of the polyoma  
PT virus  
XX  
XX Claim 1; Page 7-9; 9pp; Japanese.  
XX  
XX The amino acid sequence of the novel ETS transcription factor family

CC member - PEA3-beta. The gene was isolated from a HepG2 cell line cDNA  
CC library. The probe for the screening was prepared by reverse  
CC transcription on HepG2 mRNA followed by PCR using primers Q91770-1,  
CC to produce a probe of 170-200 bp. The screening isolated the full  
CC length sequence of the transcription factor. The gene was inserted into  
CC the expression vectors pBluescript KS and pCDV1 for expression of the  
CC protein in E.coli and COS7 cells, respectively. The ETS transcription  
CC factor has specificity for and is activated by ras. It is useful as a  
CC reagent in studies for the elucidation of the mechanism of cancer cell  
CC multiplication or polyoma virus transformation of cells.

XX  
50 Sequence 520 AA;

Query Match 11.5%; Score 94.5; DB 16; Length 520;  
Best Local Similarity 22.4%; Pred. No. 0.42;  
Matches 38; Conservative 13; Mismatches 62; Indels 57; Gaps 7;  
QY 3 HRSCHPTMTILQAPTPAPSTIPGRRGSGPEITFDPLPEPAAAPAGRPASRGHRKRS 62  
Db 162 hapaagpvqgv--gpapaphslpep-----gpqqqtfa-vprpphqpqlgmp-----kmm 207  
QY 63 RRVLYPRVVRQQLPVEEPNPAKRLFLLLTIVFCOILMAEGVP----- 106  
Db 208 pengypsegrfqrlsep-----chfpqpqpgvpgdnrpsyhrqmsepiv 252  
QY 107 --APLPPED-----APNAASLAPTPVSPVLEPFLNLTSEPSDYALD 144  
Db 253 paappppggfkqeyhdplyehgvqgmpgppahgfgsqmgikqepdrdydvd 302

Search completed: January 30, 2001, 22:05:37  
Job time: 11886 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 18:47:25 ; Search time 1450.12 Seconds  
(without alignments)  
1246.746 Million cell updates/sec

Title: US-08-799-910-9\_COPY\_211\_468  
Perfect score: 258  
Sequence: 1 GTCGGCGCCACGTCGCCAGT.....AGCAACACCGCGCGCCTTC 258

Scoring table: IDENTITY\_NUC \*  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estpl1:\*  
70: em\_estpl2:\*  
71: em\_estpl3:\*  
72: em\_estpl4:\*  
73: em\_estpl5:\*  
74: em\_estrol1:\*  
75: em\_estrol2:\*  
76: em\_estrol3:\*  
77: em\_estrol4:\*  
78: em\_estrol5:\*  
79: em\_estrol6:\*  
80: em\_estrol7:\*  
81: em\_estrol8:\*  
82: em\_estrol9:\*  
83: em\_estrol10:\*  
84: em\_estrol11:\*  
85: em\_estrol12:\*  
86: em\_estrol13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est58:\*  
105: gb\_est59:\*  
106: gb\_est60:\*  
107: gb\_est61:\*  
108: gb\_est62:\*  
109: gb\_est63:\*  
110: gb\_est64:\*  
111: gb\_est65:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*



117: em\_estp16: \*  
118: em\_estp17: \*  
119: em\_estp18: \*  
120: em\_estp19: \*  
121: em\_estp20: \*  
122: em\_estp21: \*  
123: em\_estp22: \*  
124: em\_estp23: \*  
125: em\_estp24: \*  
126: em\_estp25: \*  
127: em\_estp26: \*  
128: em\_estp27: \*  
129: em\_estp28: \*  
130: em\_estp29: \*  
131: em\_estp30: \*  
132: em\_estp31: \*  
133: em\_estp32: \*  
134: em\_estp33: \*  
135: em\_estp34: \*  
136: em\_estp35: \*  
137: em\_estp36: \*  
138: em\_estp37: \*  
139: em\_estp38: \*  
140: em\_estp39: \*  
141: em\_estp40: \*  
142: em\_estp41: \*  
143: em\_estp42: \*  
144: em\_estp43: \*  
145: em\_estp44: \*  
146: em\_estp45: \*  
147: em\_estp46: \*  
148: em\_estp47: \*  
149: em\_estp48: \*  
150: em\_estp49: \*  
151: em\_estp50: \*  
152: em\_estp51: \*  
153: em\_estp52: \*  
154: em\_estp53: \*  
155: em\_estp54: \*  
156: em\_estp55: \*  
157: em\_estp56: \*  
158: em\_estp57: \*  
159: em\_estp58: \*  
160: em\_estp59: \*  
161: em\_estp60: \*  
162: em\_estp61: \*  
163: em\_estp62: \*  
164: em\_estp63: \*  
165: em\_estp64: \*  
166: em\_estp65: \*  
167: em\_estp66: \*  
168: em\_estp67: \*  
169: em\_estp68: \*  
170: em\_estp69: \*  
171: em\_estp70: \*  
172: em\_estp71: \*  
173: em\_estp72: \*  
174: em\_estp73: \*  
175: em\_estp74: \*  
176: em\_estp75: \*  
177: em\_estp76: \*  
178: em\_estp77: \*  
179: em\_estp78: \*  
180: em\_estp79: \*  
181: em\_estp80: \*  
182: em\_estp81: \*  
183: em\_estp82: \*  
184: em\_estp83: \*  
185: em\_estp84: \*  
186: em\_estp85: \*  
187: em\_estp86: \*  
188: em\_estp87: \*  
189: em\_estp88: \*

190: gb\_gss25: \*  
191: gb\_gss26: \*  
192: gb\_gss27: \*  
193: gb\_gss28: \*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	258	100.0	411	25	A1824906	A1824906 wb03e06.x
2	258	100.0	544	107	BE384949	BE384949 601276496
3	258	100.0	951	107	BE383865	BE383865 601273192
4	256.4	99.4	765	106	R2275966	R2275966 601120727
5	254.4	98.6	762	110	BE621346	BE621346 601493736
6	249.4	96.7	916	137	BE882130	BE882130 601505227
7	245.4	95.1	603	106	BE273386	BE273386 601143493
8	244.6	94.8	463	7	AA410666	AA410666 z130q09.r
9	242	93.8	585	97	AW960290	AW960290 EST372361
10	236.8	91.8	810	109	BE563592	BE563592 601334706
11	223	86.4	447	147	W52268	W52268 zc46a10.r1
12	217.2	84.2	513	11	AA743819	AA743819 ob01d04.s
13	205.4	79.6	212	142	H80073	H80073 yv80g11.r1
14	203	78.7	1070	137	BE878459	BE878459 601487866
15	197.8	76.7	504	2	AA143404	AA143404 zo66e09.r
16	195.6	75.8	502	90	AW465331	AW465331 BP230018B
17	195.6	75.8	523	135	BE751344	BE751344 203178 MA
18	195.6	75.8	547	135	BE750116	BE750116 201294 MA
19	195.6	75.8	557	139	BF040186	BF040186 BP250005A
20	195.6	75.8	571	139	BF045952	BF045952 BP250003B
21	195.6	75.8	594	37	AV617905	AV617905 AV617905
22	192	74.4	346	38	AV683186	AV683186 AV683186
23	189.2	73.3	545	139	BF043085	BF043085 BP250026B
24	188.8	73.2	617	38	AV690660	AV690660 AV690660
25	181	70.2	184	143	N57203	N57203 yw91h09.r1
26	173.6	67.3	701	135	BE735733	BE735733 601304556
27	166	64.3	389	146	W39753	W39753 zc80b07.r1
28	163.8	63.5	174	137	BE939840	BE939840 RC6-UT001
29	160.8	62.3	597	37	AV594384	AV594384 AV594384
30	155	60.1	443	147	W47587	W47587 zc35b02.r1
31	155	60.1	806	139	BF025858	BF025858 601670012
32	154.4	59.8	628	96	AW916683	AW916683 EST347987
33	151.2	58.6	467	19	A1326484	A1326484 mq23c12.y
34	146.4	56.7	182	6	AA356940	AA356940 EST65572
35	145.8	56.5	150	89	AW367127	AW367127 MR0-HT015
36	144.8	56.1	487	139	BF023106	BF023106 ux04004.y
37	143.2	55.5	471	109	BE554106	BE554106 ur38c03.y
38	139.4	54.0	1155	135	BE786915	BE786915 601477783
39	139.2	54.0	385	147	W77628	W77628 me68a09.r1
40	139.2	54.0	460	94	AW740625	AW740625 ur02h11.y
41	137	53.1	353	142	N45391	N45391 yw97d08.r1
42	136	52.7	425	2	AA122977	AA122977 mq23c12.r
43	127.6	49.5	443	147	W71619	W71619 me40d11.r1
44	126	48.8	459	26	AI876358	AI876358 uk74a05.y
45	125.4	48.6	476	12	AA833402	AA833402 ub58b04.r

ALIGNMENTS

RESULT 1  
A1824906/C  
LOCUS A1824906 411 bp mRNA EST 16-DEC-1999  
DEFINITION wb03e06.x1 NCI-CCAP-CC6 Homo sapiens cDNA clone IMAGE:2304610 3' similar to SW:1EX1.HUMAN P46695 RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 ; mRNA sequence.  
ACCESSION A1824906  
VERSION A1824906.1 GI:5445577  
KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 411)  
 AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
 R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/bbrp/image/image.html  
 Insert Length: 557 Std Error: 0.00  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 341.  
 Location/Qualifiers  
 1. 411

FEATURES  
 source

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2304610"  
 /clone\_lib="NCI\_CGAP\_CC6"  
 /tissue\_type="pooled germ cell tumors"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Plasmid DNA from the normalized library  
 NCI-CCGAP.GC4 was prepared, and ss circles were made in  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from a pool of 5,000 clones made  
 from the same library (cloneIDs 1257096-1258631,  
 1469064-1470983, and 1475592-1476743). Subtraction by  
 Bento Soares and M. Fatima Bonaldo."  
 77 a 104 c 168 g 61 t 1 others

BASE COUNT  
 ORIGIN

Query Match 100.0%; Score 258; DB 25; Length 411;  
 Best Local Similarity 100.0%; Pred. No. 1e-55;  
 Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGCGCCAGTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
 DB 282 GTCGGCGCCAGTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 223  
 QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 120  
 DB 222 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 163  
 QY 121 CCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCCAACCCCTGTGTGCCCGCTCTCGAG 180  
 DB 162 CCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCCAACCCCTGTGTGCCCGCTCTCGAG 103  
 QY 181 CCCTTTAACTGACTTCGGAGCCCTCGGACTACGGCTCTGGACCTCAGACATTTCTCCAG 240  
 DB 102 CCCTTTAACTGACTTCGGAGCCCTCGGACTACGGCTCTGGACCTCAGACATTTCTCCAG 43  
 QY 241 CAACACCGCGCCCTTC 258  
 DB 42 CAACACCGCGCCCTTC 25

RESULT 2  
 BE384949  
 LOCUS

BE384949 544 bp mRNA EST 21-JUL-2000

## DEFINITION

601276496F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:3617543 5',  
 mRNA sequence.  
 BE384949  
 VERSION BE384949.1 GI:9330314  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 544)  
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC/DCTD/DTP  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCM284 row: p column: 24  
 High quality sequence stop: 540.  
 Location/Qualifiers  
 1. 544

FEATURES  
 source

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3617543"  
 /clone\_lib="NIH\_MGC\_20"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:  
 EcoRI; CDNA made by oligo-dr priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 98 a 225 c 131 g 90 t

BASE COUNT  
 ORIGIN

Query Match 100.0%; Score 258; DB 107; Length 544;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-55;  
 Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGCGCCAGTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
 DB 212 GTCGGCGCCAGTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 271  
 QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 120  
 DB 272 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 331  
 QY 121 CCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCCAACCCCTGTGTGCCCGCTCTCGAG 180  
 DB 332 CCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCCAACCCCTGTGTGCCCGCTCTCGAG 391  
 QY 181 CCCTTTAACTGACTTCGGAGCCCTCGGACTACGGCTCTGGACCTCAGACATTTCTCCAG 240  
 DB 392 CCCTTTAACTGACTTCGGAGCCCTCGGACTACGGCTCTGGACCTCAGACATTTCTCCAG 451  
 QY 241 CAACACCGCGCCCTTC 258  
 DB 452 CAACACCGCGCCCTTC 469

RESULT 3  
 BE383865  
 LOCUS

BE383865 951 bp mRNA EST 21-JUL-2000  
 DEFINITION 601273192F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:3614255 5',

ACCESSION BE383865  
 VERSION BE383865.1 GI:9329230  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 JOURNAL 1 (bases 1 to 931)  
 COMMENT NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC/DCTD/DTP  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCM276 row: g column: 24  
 High quality sequence stop: 747.

FEATURES  
 Location/Qualifiers  
 1..931

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3614255"  
 /clone\_lib="NIH\_MGC\_20"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: skin; Vector: pOFB7; Site\_1: XhoI; Site\_2:  
 EcoRI; CDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 210 a 324 c 253 g 164 t

Query Match 100.0%; Score 258; DB 107; Length 951;  
 Best Local Similarity 100.0%; Pred. No. 1.le-55;  
 Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCAGCTGCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60  
 |||||  
 DB 210 GTCCGGCGCAGCTGCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 269  
 |||||  
 QY 61 CTCACCATCGTCTTCTGCGAGATCCGTGATGGCTGAAGAGGGGTGTGCCGGCGCCCTGCC 120  
 |||||  
 DB 270 CTCACCATCGTCTTCTGCGAGATCCGTGATGGCTGAAGAGGGGTGTGCCGGCGCCCTGCC 329  
 |||||  
 QY 121 CCAGAGGAGCCCTTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCCGAG 180  
 |||||  
 DB 330 CCAGAGGAGCCCTTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCCGAG 389  
 |||||  
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCTAGCACTTTCCTCCAG 240  
 |||||  
 DB 390 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCTAGCACTTTCCTCCAG 449  
 |||||  
 QY 241 CAACACCGCGCCCTTC 258  
 |||||  
 DB 450 CAACACCGCGCCCTTC 467

RESULT 4  
 BE275966  
 LOCUS HE275966 765 bp mRNA  
 DEFINITION 601120727F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:2967534 5',  
 mRNA sequence.

ACCESSION BE275966  
 VERSION BE275966.1 GI:9150928  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 JOURNAL 1 (bases 1 to 755)  
 COMMENT NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC/DCTD/DTP  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCM72 row: e column: 07  
 High quality sequence start: 7  
 High quality sequence stop: 699.

FEATURES  
 Location/Qualifiers  
 1..765

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2967534"  
 /clone\_lib="NIH\_MGC\_20"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: skin; Vector: pOFB7; Site\_1: XhoI; Site\_2:  
 EcoRI; CDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 151 a 284 c 202 g 128 t

Query Match 99.4%; Score 256.4; DB 106; Length 765;  
 Best Local Similarity 99.6%; Pred. No. 2.8e-55;  
 Matches 257; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCCGGCGCAGCTGCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60  
 |||||  
 DB 257 GTCCGGCGCAGCTGCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 316  
 |||||  
 QY 61 CTCACCATCGTCTTCTGCGAGATCCGTGATGGCTGAAGAGGGGTGTGCCGGCGCCCTGCC 120  
 |||||  
 DB 317 CTCACCATCGTCTTCTGCGAGATCCGTGATGGCTGAAGAGGGGTGTGCCGGCGCCCTGCC 376  
 |||||  
 QY 121 CCAGAGGAGCCCTTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCCGAG 180  
 |||||  
 DB 377 CCAGAGGAGCCCTTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCCGAG 436  
 |||||  
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCTAGCACTTTCCTCCAG 240  
 |||||  
 DB 437 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCTAGCACTTTCCTCCAG 496  
 |||||  
 QY 241 CAACACCGCGCCCTTC 258  
 |||||  
 DB 497 CAACACCGCGCCCTTC 514

RESULT 5  
 BE275966  
 LOCUS BE275966 762 bp mRNA  
 DEFINITION 601493736F1 NIH\_MGC\_70 Homo sapiens cDNA clone IMAGE:3895736 5',  
 mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

BE521346  
BE521346.1 GI:9892286  
EST.  
human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 762)  
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

TITLE

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM674 row: h column: 09

High quality sequence stop: 613.

Location/Qualifiers

1..762

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3895736"

/clone\_lib="NIH\_MGC\_70"

/tissue\_type="epithelioid carcinoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site: 1: NotI;

Site: 2: SalI; Cloned unidirectionally. Primer: Oligo df.

Average insert size 1.1 kb. Library constructed by Life

Technologies."

BASE COUNT 155 a 288 c 193 g 126 t

ORIGIN

Query Match 98.6%; Score 254.4; DB 110; Length 762;

Best Local Similarity 99.6%; Pred. No. 8.9e-55;

Matches 255; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGGGCCAGCTGCCAGTCGAGAACCCAGCCAGCCAAAGGCTTCTTTCTGCTG 60

|||||

Db 209 GTCCGGCCCACTGCCAGTCGAGAACCCAGCCAGCCAAAGGCTTCTTTCTGCTG 268

|||||

QY 61 CTCACATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTCCGGCCCTGCT 120

|||||

Db 269 CTCACATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTCCGGCCCTGCT 328

|||||

QY 121 CCAGAGACGCCCTTAAGCCCGATCCCTGGCGCCACCCCTGTCTCCCGCTCTCGAG 180

|||||

Db 329 CCAGAGACGCCCTTAAGCCCGATCCCTGGCGCCACCCCTGTCTCCCGCTCTCGAG 388

|||||

QY 181 CCTTTAATCTGACTTGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTCTCCAG 240

|||||

Db 389 CCTTTAATCTGACTTGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTCTCCAG 448

|||||

QY 241 CAACACCCGGCGCT 256

|||||

Db 449 CAACACCCGGCGCTT 464

|||||

RESULT 6

BE882130

LOCUS BE882130 916 bp mRNA EST 27-SEP-2000

DEFINITION 60150527P2 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:3906823 5',

mRNA sequence.

ACCESSION BE882130

VERSION BE882130.1 GI:10330906

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 916)  
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

TITLE

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM703 row: f column: 08

High quality sequence stop: 659.

Location/Qualifiers

1..916

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3906823"

/clone\_lib="NIH\_MGC\_71"

/tissue\_type="leiomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: NotI;

Site: 2: SalI; Cloned unidirectionally. Primer: Oligo df.

Average insert size 2.1 kb.

BASE COUNT 187 a 327 c 268 g 134 t

ORIGIN

Query Match 96.7%; Score 249.4; DB 137; Length 916;

Best Local Similarity 99.6%; Pred. No. 1.7e-53;

Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGGGCCAGCTGCCAGTCGAGAACCCAGCCAGCCAAAGGCTTCTTCTGCTG 60

|||||

Db 180 GTCGGGCCAGCTGCCAGTCGAGAACCCAGCCAGCCAAAGGCTTCTTCTGCTG 239

|||||

QY 61 CTCACATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTCCGGCCCTGCT 120

|||||

Db 240 CTCACATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTCCGGCCCTGCT 299

|||||

QY 121 CCAGAGACGCCCTTAAGCCCGATCCCTGGCGCCACCCCTGTCTCCCGCTCTCGAG 180

|||||

Db 300 CCAGAGACGCCCTTAAGCCCGATCCCTGGCGCCACCCCTGTCTCCCGCTCTCGAG 359

|||||

QY 181 CCTTTAATCTGACTTGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTCTCCAG 240

|||||

Db 360 CCTTTAATCTGACTTGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTCTCCAG 419

|||||

QY 241 CAACACCCGGC 251

|||||

Db 420 CAACACCCGGC 430

|||||

RESULT 7

BE273386

LOCUS BE273386 603 bp mRNA EST 13-JUL-2000

DEFINITION 601143493F1 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:3507191 5',

mRNA sequence.

ACCESSION BE273386

VERSION BE273386.1 GI:9148012

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 603)

AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

# TITLE JOURNAL COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: DCTD/DRP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: L1CM186 tow: j column: 24  
High quality sequence stop: 579.  
Location/Qualifiers

## FEATURES source

1. .603  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3507191"  
/clone\_lib="NIH\_MGC\_14"  
/tissue\_type="renal cell adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/notes="Organ: kidney; Vector: pORF7; Site: 1; XhoI; Site: 2;  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(C). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
112 a 241 c 154 g 96 t

## BASE COUNT ORIGIN

Query Match 95.1%; Score 245.4; DB 106; Length 603;  
Best Local Similarity 99.2%; Pred. No. 1.7e-52;  
Matches 257; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 1 GTCCGGCGCCAGCTCCAGTCGAGAACCCAGCCAAAGGCTTCCTTCCTGCTG 60  
DB 177 GTCCGGCGCCAGCTCCAGTCGAGAACCCAGCCAAAGGCTTCCTTCCTGCTG 236  
OY 61 CTCACCATCGTTCCTGCCAGATCCTGATCGCTGAAGAGGGTGTGCCGGCGCCCTGCCCT 120  
DB 237 CTCACCATCGTTCCTGCCAGATCCTGATCGCTGAAGAGGGTGTGCCGGCGCCCTGCCCT 296  
OY 121 CCAGAGAGCGCCCTACAGCGGATCCTCGGCGCCACCCCTGTGTGCC-CCGTCCTCGA 179  
DB 297 CCAGAGAGCGCCCTACAGCGGATCCTCGGCGCCACCCCTGTGTGCCCTGCCCTCGA 356  
OY 180 GCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCGACTCAGCACTTCCTCCA 239  
DB 357 GCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCGACTCAGCACTTCCTCCA 416  
OY 240 GCAACACCGCGCCCTTC 258  
DB 417 GCAACACCGCGCCCTTC 435

## RESULT 8 AA410666

LOCUS AA410666 463 bp mRNA EST 18-MAY-1997  
DEFINITION zt30g09.r1 Soares ovary tumor NBH07 Homo sapiens cDNA clone  
IMAGE:723904.5, similar to SW:GL96\_MOUSE P46694 IMMEDIATE EARLY  
PROTEIN GLY96. ;, mRNA sequence.

## ACCESSION VERSION KEYWORDS SOURCE

AA410666  
AA410666.1 GI:2069789  
EST.  
human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 463)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

# TITLE JOURNAL COMMENT

Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, H.,  
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,  
T., Waterston, R., and Wilson, R.  
WashU-Merck EST Project 1997  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL. Contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28ml3 rev2 Et from Amerham  
High quality sequence stop: 369.  
Location/Qualifiers

## FEATURES source

1. .463  
/organism="Homo sapiens"  
/db\_xref="GDB:5935321"  
/db\_xref="taxon:9606"  
/clone="IMAGE:723904"  
/clone\_lib="Soares ovary tumor NBH07"  
/sex="Female"  
/tissue\_type="ovarian tumor"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: ovary; Vector: pMT3D (Pharmacia) with a  
modified polylinker; Site: 1; Not I; Site: 2; Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGCGCGGTTTTTTTTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pMT3 vector  
(Pharmacia). Library constructed by Bento Soares and  
M. Fatima Bonaldo."  
71 a 196 c 114 g 82 t

## BASE COUNT ORIGIN

Query Match 94.8%; Score 244.6; DB 7; Length 463;  
Best Local Similarity 98.4%; Pred. No. 2.6e-52;  
Matches 247; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GTCCGGCGCCAGCTCCAGTCGAGAACCCAGCCAAAGGCTTCCTTCCTGCTG 60  
DB 213 GTCCGGCGCCAGCTCCAGTCGAGAACCCAGCCAAAGGCTTCCTTCCTGCTG 272  
OY 61 CTCACCATCGTTCCTGCCAGATCCTGATCGCTGAAGAGGGTGTGCCGGCGCCCTGCCCT 120  
DB 273 CTCACCATCGTTCCTGCCAGATCCTGATCGCTGAAGAGGGTGTGCCGGCGCCCTGCCCT 332  
OY 121 CCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGTGCCCGCTCCGAG 180  
DB 333 CCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGTGCCCGCTCCGAG 392  
OY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACTCAGCACTTCCTCCAG 240  
DB 393 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACTCAGCACTTCCTCCAG 452  
OY 241 CAACACCGCGC 251  
DB 453 CAACACCGCGC 463

## RESULT 9 AW960290

LOCUS AW960290 585 bp mRNA EST 01-JUN-2000  
DEFINITION EST372361 MAGE resequences, MAGF Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW960290  
VERSION AW960290.1 GI:8149974  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 585)  
 AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and Quackenbush, J.

TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

JOURNAL: Unpublished (2000)

COMMENT Contact: John Quackenbush  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528

Fax: 301 838 0208

Email: johnq@tigr.org

Plate: 144

Seq primer: Reversal

Location/Qualifiers

1. .585

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="MAGE resequences, MAGF"

/note="Vector: pBluescriptSKm"

134 a 190 c 148 g 113 t

BASE COUNT

ORIGIN

Query Match 93.8%; Score 242; DB 97; Length 585;  
 Best Local Similarity 96.1%; Pred. No. 1.2e-51;  
 Matches 248; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GTCGGCGCCAGCTGCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTCTTTCTGCTG 60  
 Db 19 GCGCGCGCCAGCTGCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTCTTTCTGCTG 78  
 QY 61 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCCT 120  
 Db 79 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCCT 138  
 QY 121 CCAGAGACGCCCTTACGCCGATCCCTGCGCGCCACCCCTGTGTCGCCCGCTCGAG 180  
 Db 139 CCAGAGACGCCCTTACGCCGATCCCTGCGCGCCACCCCTGTGTCGCCCGCTCGAG 198  
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCTCCAG 240  
 Db 199 CCCTTTAATCTGACTTGTGAGCCCTCGGACTACACTCTGGACCTCAGCACTTCTCTCCAG 258  
 QY 241 CACACCCCGCGCCTTC 258  
 Db 259 CAACACCCCGCGCCTTC 276

RESULT 10

BE563592

LOCUS

DEFINITION BE563592 810 bp mRNA EST 15-AUG-2000  
 601334706f1 NIH\_MGC\_39 Homo sapiens cDNA clone IMAGE:3688825 5', mRNA sequence.

ACCESSION BE563592

VERSION BE563592.1

KEYWORDS GI:9807312

SOURCE EST.

ORGANISM human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 810)

NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLC382 row: k column: 02

High quality sequence stop: 692.

FEATURES

source

1. .810

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3688825"

/clone\_lib="NIH\_MGC\_39"

/tissue\_type="adenocarcinoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI;

Site: 2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCAGAG(G). Library constructed

by Ling Hong in the laboratory of Gerald M. Rubin

(University of California, Berkeley) using ZAP-cDNA

synthesis kit (Stratagene) and Superscript II RT (Life

Technologies).

176 a 284 c 205 g 145 t

BASE COUNT

ORIGIN

Query Match 91.8%; Score 236.8; DB 109; Length 810;  
 Best Local Similarity 98.8%; Pred. No. 2.6e-50;  
 Matches 249; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GTCGGCGCCAGCTGCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTCTTTCTGCTG 60  
 Db 207 GTCGGCGCCAGCTGCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTCTTTCTGCTG 266  
 QY 61 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCCT 120  
 Db 267 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCCT 326  
 QY 121 CCAGAGACGCCCTTACGCCGATCCCTTGGCGCCACCCCTGTGTCGCCCGCTCGA 179  
 Db 327 CCAGAGACGCCCTTACGCCGATCCCTTGGCGCCACCCCTGTGTCGCCCGCTCGA 386  
 QY 180 GCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGACCTCAGCACTTTCCTCCA 239  
 Db 387 GCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGACCTCAGCACTTTCCTCCA 446  
 QY 240 GCAACCCCGCGC 251  
 Db 447 GCAACCCCGCGAC 458

RESULT 11

W52268

LOCUS

DEFINITION W52268 447 bp mRNA EST 11-OCT-1996  
 zc46a10.r1 Soares\_senescent\_fibroblasts\_NHHSF Homo sapiens cDNA clone IMAGE:325338 5' similar to SW:GL96\_MOUSE I46694 IMMEDIATE EARLY PROTEIN GLY96. [1]; mRNA sequence.

ACCESSION W52268

VERSION W52268.1

KEYWORDS GI:1349380

SOURCE EST.

ORGANISM human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 447)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine



4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 192

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: M13RP1

High quality sequence stop: 192.

FEATURES

Location/Qualifiers  
1..212  
/organism="Homo sapiens"  
/db\_xref="GDB:3866798"  
/db\_xref="taxon:9606"  
/clone="IMAGE:249092"  
/sex="Male"  
/sex="Male"  
/tissue\_type="melanocyte"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT7T3D (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TCTTACCAATCTGAAGTGGAGCGCGCGAGTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT7T3 vector  
(Pharmacia). Library constructed by Bento Soares and  
M. Fallima Bonaldo. RNA from normal foreskin melanocytes  
(FS374) was kindly provided by Dr. Anthony P. Albino."

34 a 81 c 49 g 44 t

4 others

0; Gaps 0;

Indels 5; Mismatches 0;

Length 212;

Score 205.4; DB 142;

Best Local Similarity 97.6%;

Conservative 0;

Query Match

79.6%;

Score 205.4; DB 142;

Length 212;

Best Local Similarity 97.6%;

Conservative 0;

Query Match

79.6%;

Score 205.4; DB 142;

Length 212;

Best Local Similarity 97.6%;

Conservative 0;

Query Match

79.6%;

Score 205.4; DB 142;

Length 212;

Best Local Similarity 97.6%;

Conservative 0;

Query Match

79.6%;

Score 205.4; DB 142;

Length 212;

Best Local Similarity 97.6%;

Conservative 0;

Query Match

79.6%;

Score 205.4; DB 142;

Length 212;

Best Local Similarity 97.6%;

Conservative 0;

Query Match

79.6%;

Score 205.4; DB 142;

Length 212;

Best Local Similarity 97.6%;

Conservative 0;

Query Match

79.6%;

Score 205.4; DB 142;

Length 212;

Best Local Similarity 97.6%;

Conservative 0;

Query Match

79.6%;

Score 205.4; DB 142;

Length 212;

Best Local Similarity 97.6%;

Conservative 0;

Email: Robert\_Strausberg@nih.gov

Tissue Procurement: DCTD/DRP/Gazdar

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM660 row: c column: 17

High quality sequence stop: 645.

FEATURES

Location/Qualifiers  
1..1070  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3890248"  
/clone\_lib="NIH\_MGC\_69"  
/tissue\_type="large cell carcinoma, undifferentiated"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.1 kb. Library constructed by Life  
Technologies."

BASE COUNT 276 a 324 c 302 g 168 t

ORIGIN

Query Match 78.7%; Score 203; DB 137; Length 1070;

Best Local Similarity 98.1%; Pred. No. 1e-41; 0; Indels 5; Gaps 5;

Matches 258; Conservative 0; Mismatches 0; Indels 5; Gaps 5;

QY 1 GTCCGCGCCAG-CTGCCAGTCGAGGACCGAACCCAGCCAA-AAGGCITCTCTTTCTGC 58

Db 63 GTCCGCGCCAGTCGCCAGTCGAGGACCGAACCCAGCCAA-AAGGCITCTCTTTCTGC 122

QY 59 TGC-TCACCATCTCTTCTGCCAGATCCTGATCGCTGAGAGGGTGTGCGGCGCCCTG 117

Db 123 TCATCACCATCTCTTCTGCCAGATCCTGATCGCTGAGAGGGTGTGCGGCGCCCTG 182

QY 118 CTCCAGAGGAC-GCCCTTAACGCCGATCCCTGCG-CCACCCCTGTGTCCCGCTCC 175

Db 183 CTCCAGAGGACGTCCTTAACGCCGATCCCTGCG-CCACCCCTGTGTCCCGCTCC 242

QY 176 TCGAGCCCTTAATCTGACITCGAGCCCTCGGACCTCGGACCTCGGACCTCGGACCTTC 235

Db 243 TCGAGCCCTTAATCTGACITCGAGCCCTCGGACCTCGGACCTCGGACCTTC 302

QY 236 TCCAGCAACACCGCGCGCTTC 258

Db 303 TCCAGCAACACCGCGCGCTTC 325

RESULT 15

AA143404

LOCUS

DEFINITION

z066e09.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone

IMAGE:591880 5' similar to SW:GL96\_MOUSE P46694 IMMEDIATE EARLY

PROTEIN GLY96. ;, mRNA sequence.

AA143404

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

1 (bases 1 to 504)

Hillier, L., Allen, M., Bowles, L., Dubouque, T., Geisel, G., Jost, S.,

Krisman, D., Kucaba, T., Lacy, M., Lee, N., Lennon, G., Marra, M., Martin

, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,

White, Y., Wyllie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

Contact: Wilson RK

Washington University School of Medicine

REFERENCE:

AUTHORS

1 (bases 1 to 1070)

NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

RESULT 14

BE878459

LOCUS

DEFINITION

BE878459 1070 bp mRNA

EST

27-SEP-2000

IMAGE:3890248 5'

mRNA sequence.

BE878459

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

1 (bases 1 to 1070)

NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1070)

NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550



4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LILNL; contact the IMAGE Consortium (info@image.lilnl.gov) for further information.

Insert Length: 1695 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 392.

FEATURES

source

1..504

/organism="Homo sapiens"

/db\_xref="GDB:462326"

/db\_xref="taxon:9606"

/clone="IMAGE:591880"

/clone\_lib="Strategene pancreas (#937208)"

/lab\_host="SOUR cells (kanamycin resistant)"

/note="Organ: pancreas; Vector: pBluescript SK-; Site.1:

ECORI; Site.2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. Pancreatic adenocarcinoma cell line. Average

insert size: 1.0 Kb; Uni-ZAP XR vector; -5' adaptor

sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTT 3'

BASE COUNT  
ORIGIN

80 a 202 c 121 g 92 t 9 others

Query Match 76.7%; Score 197.8; DB 2; Length 504;  
Best Local Similarity 90.7%; Pred. No. 2e-40;  
Matches 233; Conservative 0; Mismatches 19; Indels 5; Gaps 2;

QY 1 GTCCGGCGCAGTGCAGTCGAGGAACCGAACCC-AGCCAAAGGCTTCTCTTCTGCT 59  
DB 225 GTCCGGCGCAGTGCAGTCGAGGAACCGAACCCAAAGGCTTCTCTTCTGCT 284  
QY 60 GCTACCATCGTCTTGTGCCAGATCCTGATGGCTGAAGAGGCTGTGCGGCCCTGCC 119  
DB 285 GCTACCATCGTCTTGTGCCAGATCCTGATGGCTGAAGAGGCTGTGCGGCCCTGCC 344  
QY 120 TCCAGAGAGCGCCCTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCCG 179  
DB 345 TCCAGAGAGCGCCCTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCCG 404  
QY 180 GCCCTTTAATCTGACTTCGGAGCGCCTCGG-...ACTACGCTCTGGACCTCAGCAGTTCC 235  
DB 405 GCNCTTTAATCTGACTTCGGAGCGCCTTGGAACTANGCTTCTTGGAACTTAAGCAGTTCC 464  
QY 236 TCCAGACACACCCGCC 252  
DB 465 TCCAGACACACCCGCC 481

Search completed: January 30, 2001, 18:47:28  
Job time: 28459 sec

---

CC by RNA polymerase II, the RNA polymerase which transcribes messenger  
CC RNA. RNA polymerase II transcription proceeds *in vitro* upon addition  
CC of several nuclear fractions designated TFIIA, B, D, F, H, I and J  
CC to RNA polymerase II holoenzyme. Fraction TFIIF has been shown to  
CC contain a TAP and other TAPs. Purification of TFIIF and separation of  
CC its components reveals 7 proteins ranging in size from 30-250 kD.  
CC Serum raised against the TFIIF fraction allowed cloning of the corresp.  
CC genes from lambda-gt10 expression libraries.  
XX  
SQ Sequence 1120 BP; 307 A; 291 C; 312 G; 210 T; 0 other;

Query Match	13.0%;	Score 33.6;	DB 17;	Length 1120;
Best Local Similarity	51.0%;	Pred. No. 2.2;		
Matches 104;	Conservative 0;	Mismatches 99;	Indels 1;	Gaps
QY	34	CCAGCAGAAAGCTTCTCTTCTGCTGCTCCACATCGTCTCTGCGACATCTCGATGCGT	93	
Db	866	CCATCCACAGCCAGATGATCTCTCTCTCTCTCTGCGTGGACATTCCTGCT-CCC	808	
QY	94	GAGAGGGTGTGCCGGCGCCCTCGCTCCAGAGGACGCCCTTAAGCGCGCATCCCTGGCG	153	
Db	807	GAACTGGATGCCCCAGCATGTGTGCCACGGCGCGCATCGCTGCCACCTCCATCTTG	748	
QY	154	CCCACCCCTGTCTGCCCGTCTCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGACATAC	213	
Db	747	AGATCTCCCGCGCGCGTGTCTCGCTCTTTAACCCTCTTGACCAACCGCCGCCCCAGAT	688	
QY	214	GTCTGGACCTCAGCACTTTCCCTC	237	
Db	687	CGCGCGAGCTGCCACCGTTTTTC	664	

RESULT 15

ID T79601 standard; cDNA; 1120 BP.

AC T79601;

D11 08-OCT-1997 (first entry)

DE TATA-binding protein associated factor, dTAFFI40 cDNA.

...  
KW TATA-binding protein associated factor; TAF; nuclear protein;  
KW RNA polymerase transcription; TATA-binding protein; TBP;  
KW initiation; ds.

OS *Drosophila* sp.

	Key	Location/Qualifiers
CDS		80..916
		/*tag= a

PN US5637686-A.

PD 10-JUN-1997.

28-JAN-1993; 93US-0013412.

28-JAN-1994: 94US-0188582.

93US-0013412;  
93US-0087119;

XX  
03 MAR 1990, 3003-0040/13.XX  
 (REC ) UNIV CALIFORNIA.

COMAI L., DYNIAK BD., HOEY T., RUPPERT S., TANASE N.,  
PI  
TILAN B., WANG E. WEINZIERL ROJ:

WPI: 1997-319113/29

DR P-PSDB; W25025.

Nucleic acids encoding human TATA-binding protein associated factor

PT	(TAF) peptide(s) - for production of recombinant peptide(s), used
PT	for modulating transcription of TAFs
XX	
XX	Example 1; Column 61-64; 86pp; English.
XX	
CC	T79601 encodes Drosophila TATA-binding protein associated factor
CC	(TAF) polypeptide, dTAFII40 (mol. weight 40kD). TAF peptides derived
CC	from dTAFII30 alpha, dTAFII30 beta, dTAFII40, dTAFII60, dTAFII80,
CC	dTAFII110, dTAFII150, and dTAFII250, their human equivalents and
CC	nucleic acids encoding them, are used to modulate transcription,
CC	including transcription initiation. TAFs are nuclear proteins involved
CC	in RNA polymerase I, II and III transcription. The peptides act by
CC	binding to a different TAF, an activator, or TBP (TATA-binding protein)
CC	or competitively inhibiting association of a TAF domain with another
CC	compound, typically a protein like TBP or another TAF, an activator,
CC	or DNA.
XX	
XX	Sequence 1120 BP; 307 A; 291 C; 312 G; 210 T; 0 other:
XX	

Query Match	13.0%	Score 33.6	DB 18	Length 1120
Best Local Similarity	51.0%	Pred. NO. 2.2		
Matches 104	Conservative 0	Mismatches 99	Indels 1	Gaps 1
Qy	34	CCAGCCAAAAGGCTTCCTTTCTGCTGCTCACCATGCTCTTCGCCAGATCCTCATGGCT	93	
Db	866	CCATCCAACGCCAGATGATTCCTCCTCTCTCTCCGCTGGCACTTCCTGCT-CCC	808	
Qy	94	GAGAGGGTGTGGCGGCGCCCTGCCCTCCAGAGACGCCCTTAACGCCGATCCCTCGCG	153	
Db	807	GAACCGGATCGGCAGCGATGCTGCCACGCCGCCGATCGCTGCACCTCCATCTTG	748	
Qy	154	CCACGCCCTGTGTCGCCCGCTCTCGAGCCCTTTAAATCTGACTCGGAGCCCTCGGACTAC	213	
Db	747	AGATCTCCGCGCGCGGTGCTCTGCTCTTAACCTCTGACACCGCGCCGCCACAGAT	688	
Qy	214	GCTCTGGACCTCAGCACTTTCTTC	237	
Db	687	CCGCCGAGCTGCCACCGTTTC	664	

Search completed: January 30, 2001, 21:29:32  
Job time: 19258 sec

CC surface molecule 4F2. Antisense DNA or RNA, antibodies, antibody  
 CC fragments, and peptide and non-peptide antagonists to LAT1 are useful  
 CC as anticancer agents. The present sequence encodes human LAT1, which is  
 CC specifically claimed in the present invention.

XX Sequence 4539 BP; 765 A; 1450 C; 1297 G; 998 T; 29 other;

Query Match 13.1%; Score 33.8; DB 21; Length 4539;  
 Best Local Similarity 53.0%; Pred. NO. 2.6;  
 Matches 62; Conservative 2; Mismatches 53; Indels 0; Gaps 0;

QY 8 GCCAGCTGCCAGTCGAGACACCAACCCCAAGGCTTCTTCTGCTCTCAACA 67  
 III | III | III | III | III | III | III | III | III | III |  
 Db 556 gccggtyccggaggaggaagcgaagctgtgctggytntdygtgctgctcaagg 615  
 III | III | III | III | III | III | III | III | III | III |  
 QY 68 TCGTCTTCTGCCAGATCCTCATGCTGAAGAGCTGTGCCGCCGCCCTGCCCTCCAG 124  
 III | III | III | III | III | III | III | III | III | III |  
 Db 616 ccgtgaactgctacagcgtgaaggcngcnaacncgggtccaggatgcttgcgcgcg 672

## RESULT 13

Q70728/c  
 ID 070728 standard; cDNA; 1120 BP.

XX AC Q70728;

XX DT 23-MAR-1995 (first entry)

XX DE TATA-binding protein-associated factor dTAFLI40 cDNA.

XX KW TATA-binding protein associated factor; dTAFLI40; ss; screening;  
 diagnostic; therapeutic; gene transcription regulation.

XX OS Drosophila.

XX FH Key Location/Qualifiers  
 CDS 80..913  
 FT /\*tag= a

XX XX W09417087-A.

XX PN 04-AUG-1994.

XX PD 28-JAN-1994; 94WO-US01114.

XX PF 28-JAN-1993; 93US-0013412.

XX PR 30-JUN-1993; 93US-0087119.

XX PA (REGC ) UNIV CALIFORNIA.

XX PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;  
 TJian R, Wang E, Weinzierl ROJ;

XX DR WPI; 1994-264019/32.  
 P-PSDB; R56490.

XX PT TATA-binding protein associated protein factors - and  
 corresponding nucleotide sequence and deriv. antibodies, useful  
 in screening, diagnostics and therapeutics

XX PS Disclosure; Page 89-92; 180pp; English.

XX CC The TATA-binding protein associated factor dTAFLI40 (including  
 specific antibodies and fusion products) are used in drug screening,  
 CC diagnostics and therapeutics. They are used in the development of  
 CC specific biochemical assays for screening compounds that agonise or  
 CC antagonise selected transcription factors involved in regulating  
 CC gene expression associated with human pathology.

XX SQ Sequence 1120 BP; 307 A; 292 C; 311 G; 210 T; 0 other;

## Query Match

Best Local Similarity 13.0%; Score 33.6; DB 15; Length 1120;  
 Matches 104; Conservative 0; Mismatches 99; Indels 1; Gaps 1;

QY 34 CCAGCAAAAGGCTTCTTCTGCTGCTCACCATGCTCTTCCGACATCCCTGATGGCT 93  
 III | III | III | III | III | III | III | III | III | III |  
 Db 866 CCACTCCAAAGCCAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 808  
 QY 94 GAAGAGGGTCCCGCGCCCTGCTCCAGAGAGAGGCGCCCTAACGCCGCATCCCTGGCG 153  
 III | III | III | III | III | III | III | III | III | III |  
 Db 807 GAACCGGATGCCCGACCGATGCTGCCCGCGCGCGCATCTGTCCACCTCCATCTTG 748  
 QY 154 CCCACCCCTGTCTCCCGCTCTCGAATCTTAATCTGACTTCGGAGCGCTCGGACTAC 213  
 III | III | III | III | III | III | III | III | III | III |  
 Db 747 ACATCTCCCGCGCGCGCTGCTCTGCTCTTAACCTCTGACCAACGCCGCCGCCCAT 688  
 QY 214 GCTCTGGACCTCAGCACTTTCCTC 237  
 III | III | III | III | III | III | III | III | III | III |  
 Db 687 CCGCGGAGCTGCCACCGTTTTC 664

## RESULT 14

T42214/c  
 ID T42214 standard; cDNA; 1120 BP.

XX AC T42214;

XX DT 27-JAN-1997 (first entry)

XX DE Drosophila TATA-binding protein associated factor dTAFLI60 gene.

XX KW Drosophila; TATA-binding protein; TBP associated factor; TFIID;  
 RNA polymerase II; transcription; messenger RNA; nuclear fraction;  
 holoenzyme; lambda-gt11; expression library; ds.

XX OS Drosophila melanogaster.

XX FH Key Location/Qualifiers  
 CDS 80..916  
 FT /\*tag= a  
 FT /product= Drosophila TAFI160

XX XX US5534410-A.

XX PD 09-JUL-1996.

XX PF 28-JAN-1993; 93US-0013412.

XX PR 28-JAN-1994; 94US-0188582;

XX PR 28-JAN-1993; 93US-0013412.

XX PR 30-JUN-1993; 93US-0087119.

XX PA (REGC ) UNIV CALIFORNIA.

XX PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;  
 TJian R, Wang E, Weinzierl ROJ;

XX DR WPI; 1996-333245/33.  
 P-PSDB; T42214.

XX PT Screen for cpds. that bind human TATA-binding protein associated  
 factor - by testing ability to bind to polypeptide fragments of the  
 PT factor, useful as (ant)agonists of transcription factors involved in  
 PT disease.

XX PS Examples; Column 61-64; 86pp; English.

XX CC This is the nucleotide sequence encoding the Drosophila TATA-binding  
 CC protein (TBP) associated factor (TAF) designated TAFI160. The protein  
 CC is a component of the TFIID fraction required for reconstituting RNA  
 CC polymerase II in vitro transcription activity. The encoded protein  
 CC has an estimated mol. wt. of 60 kD by SDS-PAGE.  
 CC The invention relates to purified proteins involved in transcription

CC attack, neurological disorders, e.g. neuroblastoma, glioblastoma or  
 CC cancers, or to promote apoptosis in cells, or treat refractory  
 CC epilepsy. They can also be used for screening test substances for  
 CC IBI biological activity. IBI producing cells can be used  
 CC therapeutically to produce IBI in a subject.

XX  
 SQ Sequence 2953 BP; 620 A; 906 C; 840 G; 587 T; 0 other;

Query Match 13.5%; Score 34.8; DB 19; Length 2953;  
 Best Local Similarity 62.8%; Pred. No. 1.3;  
 Matches 54; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 98 AGGATGTCGGCGCCCTGCTCCAGAGGAGCCCTTAACGCCGATCCCTGGCGCCCA 157  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 2539 AGGATGTCAGCGCCCTGCTCCAGGAGGACCCCTTTGTGGCCAGATGCTGACCCCA 2480  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 158 CCCCTGTGTCCCGCTCTCGAGCCC 183  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 2479 CCCAGCCCTGCCCTGCCAGCCCC 2454

RESULT 11  
 Q30849/c  
 ID Q30849 standard; cDNA; 3690 BP.

XX AC Q30849;

DT 24-MAR-1993 (first entry)

XX Type III procollagen (prior art).

DE Mutation; pro-alpha(I(III)); primer; PCR; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FT CDS 103..3690  
 FT FT /\*tag= a  
 FT FT /label= Type\_III\_procollagen

PN W09219754-A.

XX 12-NOV-1992.

XX 08-MAY-1992; 92WO-US03866.

XX 08-MAY-1991; 91US-0696607.

XX (UYJE-) UNIV JEFFERSON THOMAS.

XX Kuivaniemi SH, Prockop DJ, Tromp GC;

XX WPI; 1992-398878/48.

XX P-PSDB; R28916.

XX Kit for detecting genetic pre-disposition for vascular aneurysms  
 PT - contains primer to amplify portions of Type III procollagen DNA  
 PT and detects mutation in standard procollagen DNA

XX Disclosure; Fig 1A-F; 44pp; English.

XX Example 1 describes the determination of the presence of a mutation  
 CC in the pro-alpha(I(III)) gene. Primers used in PCR are given in  
 CC Q30834-48.

XX Sequence 3690 BP; 771 A; 1031 C; 1152 G; 736 T; 0 other;

Query Match 13.3%; Score 34.2; DB 13; Length 3690;  
 Best Local Similarity 46.2%; Pred. No. 1.9;  
 Matches 114; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy 1 GTCGCGCGCCAGCTGCCAGTCGAGAACCCAGCCAAAGGCTTCCTCTTCCTG 60  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1201 GTCAGGGGACCACTTTGAACAGGAGACCTCGAGGTCCCACTTCACCTTAGCACCAG 1142  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 61 CTCACCATGCTCTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCGGGCCCTGCTCCCT 120  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1141 GGGATCCAGGAATGCGGCAGTTCACAGGAGGACCCAGGAGGGGTGGTTCAGCATCACTGC 1082  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 121 CCAGAGACGCGCCCTAACGCGCATCCCTGCGGCCACCCCTGCTCCCGCTCCCTGAG 180  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1081 CTCGAGACCGTCACTATACCCCGACCACTCGAGGCCCGAGGAAGTCTTGGCGGTGCTCGCT 1022  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 181 CCCTTTAATCTGACTTCGAGGCCCTCGGACTACGCTCTGACCTTCAGACTTCCTCCAG 240  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1021 CACGAGAGCCCGTCTTGGAGCATGGTTCGAGGAGCTGGATTTCCTCCCTTGAGACCAT 962  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 241 CAACACC 247  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 961 TTTCACC 955

RESULT 12  
 A08372

ID A08372 standard; cDNA; 4539 BP.

XX AC A08372;

DT 12-JUL-2000 (first entry)

DE Human L-type amino acid transporter 1 nucleotide sequence SEQ ID NO:1.  
 XX L-type amino acid transporter 1; LAT1; 4F2hc; 4F2 heavy chain;  
 KW neutral amino acid transporter; tumour; cancer; proliferation;  
 KW cell membrane surface 4F2 molecule; anticancer; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FT 5'UTR 1..65  
 FT CDS /\*tag= a  
 FT CDS 66..1589  
 FT /\*tag= b  
 FT /\*product= "LAT1"  
 FT 3'UTR 1590..4474  
 FT /\*tag= c

XX WO200014228-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-JP04789.

XX 03-SEP-1998; 98JP-0249993.

XX 02-SEP-1999; 99JP-0248546.

XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX Endou H, Kanai Y;

XX WPI; 2000-256979/22.

XX P-PSDB; Y82488.

XX Neutral amino acid transporter protein which conjugates with cell  
 PT membrane protein 4F2 and operates independently of sodium ions is  
 PT useful for screening potential cancer proliferation inhibitors -  
 XX Claim 9; Page 125-132; 189pp; Japanese.

XX The present invention describes a cell surface protein which is an  
 CC L-type amino acid transporter-1 (LAT1), which mediates the transport  
 CC of neutral amino acids, leucine, isoleucine, phenylalanine, methionine,  
 CC tyrosine, tryptophan, valine and histidine, into the cell independently  
 CC of sodium ions. The LAT1 protein conjugates with the cell membrane





PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

PI Belt JA, Crawford CR, Patel DH;

XX WPI: 1998-594576/50.

XX New equilibrative nucleoside transport protein insensitive to  
PT nitrobenzylthio-purine riboside - useful for, e.g. identifying  
PT specific therapeutic nucleoside analogues and in gene therapy to  
PT protect transduced cells against ablative chemotherapy

XX Example 3; Pages 76-79; 114pp; English.

XX This represents the 5' UTR (untranslated region) of the DNA that  
CC encodes an equilibrative nucleoside transport protein (iENTP), which  
CC is insensitive to nitrobenzylmercaptopyrine riboside (NBMPR). Cells  
CC transformed with a construct containing the iENTP nucleic acid can be  
CC used to produce the protein recombinantly. iENTP is used to identify  
CC specific ligands (particularly antiviral and antitumour nucleoside  
CC analogues that are preferentially transported into cells) and to raise  
CC antibodies. Cells in which iENTP provides all available transport  
CC activity are used; (a) to identify permeants of iENTP and (b) to screen  
CC specific inhibitors of iENTP (potential drugs). Fragments of the iENTP  
CC nucleic acid are used, as probes, primers, antisense molecules, and  
CC ribozymes for therapy or diagnosis, and knockout mice in which both  
CC alleles encoding iENTP contain an inactivating defect are also useful for  
CC drug screening. Cells that have been transduced with iENTP nucleic acid  
CC ex vivo are used particularly, for cancer chemotherapy. Vectors in which  
CC the iENTP gene is linked to a heterologous gene (e.g. encoding adenosine  
CC deaminase and factor VIII) can be used for gene therapy.

XX Sequence 6354 BP; 1322 A; 1797 C; 1873 G; 1355 T; 7 other;

Query Match 15.0%; Score 38.6; DB 19; Length 6354;  
Best Local Similarity 65.9%; Pred. NO. 0.14; Indels 0; Gaps 0;  
Matches 56; Conservative 0; Mismatches 29;

QY 100 GTGTGCGCGCGCCCTGCTCCAGAGAGCGCCCTTAAGCGCGCATCCTCGGCCAC 159

DB 2461 ggtgagagcgctgcctgcctgcctctgcgagagcagcttcattgagccctcc 2520

QY 160 CTGTGTCCCGCCCTCTCGAGCCCT 184.

DB 2521 cctgcgccccctgcctccagccct 2545

RESULT 6

ID 251683

XX 251683 standard; cDNA; 4228 BP.

XX AC 251683;

XX DT 04-JUL-2000 (first entry)

XX DE Human cyclic nucleotide-associated protein-2 (CNAP-2) cDNA.

XX Cyclic nucleotide-associated protein-2; CNAP-2; human; cytosolic;  
KW anti-arteriosclerotic; hepatotropic; anti-leukaemic; anti-inflammatory;  
KW immunomodulatory; anti-asthmatic; anti-anaemic; anti-diabetic; diagnosis;  
KW anti-sclerotic; dermatological; neuroprotective; anti-epileptic; cancer;  
KW anti-Alzheimer's; anti-Parkinsonian; cerebroprotective; ophthalmological;  
KW anti-infertility; anti-allergic; vasotropic; immunosuppressive;  
KW hypotensive; gene therapy; prevention; treatment; arteriosclerosis;  
KW cell proliferative disorder; autoimmune/inflammatory; diabetes mellitus;  
KW neurological; vision; reproductive; smooth muscle; ss.

XX Homo sapiens.

XX Location/Qualifiers

XX 31..4014

XX /tag= a

XX /product= "Human CNAP-2 protein"

FT /note= "Shares 24% identity to Aquifex pyrophilus  
FT esterase 28LC"  
FT 31..132  
FT /tag= b  
FT 133..4011  
FT /tag= c  
FT /product= "Mature CNAP-2 protein"  
FT 136..165  
FT /tag= d  
FT /bound\_moiety= "Primer or Probe"  
FT /note= "Useful for amplification or hybridisation  
FT techniques"

XX WO200014248-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-US20287.

XX 04-SEP-1998; 98US-0148904.

XX (INCY-) INCYTE PHARM INC.

PI Hillman JL, Yue H, Guegler KJ, Corley NC, Patterson C, Tang YT;  
XX WPI: 2000-256994/22.

XX P-PSDB; Y70474.

XX Isolated cyclic nucleotide associated proteins useful for preventing,  
PT diagnosing and treating cell proliferative, autoimmune/inflammatory,  
PT neurological, vision, reproductive and smooth muscle disorders -

XX Claim 9; Page 70-71; 78pp; English.

XX The present sequence is the cDNA encoding human cyclic nucleotide  
CC associated protein-2 (CNAP-2), identified in Incyte clone 3149674,  
CC that is isolated from ADRENON04 cDNA library. It is expressed in  
CC nervous, reproductive, cardiovascular and haematopoietic/immune tissues.  
CC CNAP sequences may be used for prevention, treatment and diagnosis of  
CC diseases associated with altered CNAP expression such as, cell  
CC proliferative disorders (e.g. arteriosclerosis, cirrhosis, leukaemia,  
CC lymphoma and cancer of the breast, prostate, lung and brain), autoimmune/  
CC inflammatory disorders (e.g. asthma, anaemia, diabetes mellitus, multiple  
CC sclerosis and psoriasis), neurological disorders (e.g. epilepsy,  
CC Alzheimer's/Parkinson's disease and strokes), vision disorders (e.g.  
CC conjunctivitis, glaucoma, cataracts and retinitis pigmentosa),  
CC reproductive disorders (e.g. infertility, uterine fibroids, ectopic  
CC pregnancies and impotence) and smooth muscle disorders (e.g. angina,  
CC anaphylactic shock, Kearns-Sayre syndrome and hypertension). It can also  
CC be used for gene therapy.

XX Sequence 4228 BP; 793 A; 1365 C; 1324 G; 746 T; 0 other;

Query Match 14.8%; Score 38.2; DB 21; Length 4228;  
Best Local Similarity 54.7%; Pred. NO. 0.17;  
Matches 76; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 75 CTGCGAGATCCTGATGGGTGAGAGGGGTGTCGGCGGCCCTCCAGAGAGCGCCCC 134

DB 1185 ctccgacttcgacatggcctatgagcgtggccggtatctccgtccctgcaggaagggc 1244

QY 135 TAACGCCGATCCTCGGCCGCCACCCCTGTGTGCCCGCTCCGAGAGCCCTTTAATCTGAC 194

DB 1245 ctccgggggggtccctcggcagcccccgcgtcggacccccactcaggaagcctcgtgagcagcc 1304

QY 195 TTCGAGCCCTCGGACTAC 213

DB 1305 ggcaggcgctgtgaatc 1323

RESULT 7

V69706



DT 15-FEB-1999 (first entry)  
XX EST clone BY66.  
XX  
XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
KW tissue growth; activin; inhibitor; chemotaxis; chemokinesis; haemostatic;  
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
XX gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
XX  
XX WO9845436-A2.  
XX  
XX 15-OCT-1998.  
XX  
XX 10-APR-1998; 98WO-US06955.  
XX  
XX 10-APR-1997; 97US-0838821.  
XX  
XX (GEM) GENETICS INST INC.  
XX  
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
PI Racie LA, Spaulding V, Treacy M;  
XX  
XX WPI; 1999-070077/06.  
XX  
XX New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries.  
XX  
XX Claim 1; Page 164; 618pp; English.  
XX  
XX The present sequence represents a human expressed sequence tag (EST).  
CC The polynucleotide, which is a secreted EST, and the encoded protein  
CC are predicted to have useful biological activities which would make  
CC them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals, although no supporting data is  
CC given. Suggested activities include nutritional activity, immune  
CC stimulating or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
XX  
XX Sequence 193 BP; 35 A; 39 C; 68 G; 51 T; 0 other;  
SQ

Query Match 32.6%; Score 84; DB 20; Length 193;  
Best Local Similarity 100.0%; Pred. No. 4.2e-14;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 175 CTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTC 234  
DB 193 CTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTC 134  
QY 235 CTCGAGCAACACCCGCCCTTC 258  
DB 133 CTCGAGCAACACCCGCCCTTC 110  
RESULT 4  
QY7534  
ID Q77534 standard; DNA; 297 BP.  
XX  
XX Q77534;  
XX  
XX 23-SEP-1994 (first entry)  
XX  
XX Human genome fragment.  
XX  
XX Brain; placenta; bone marrow; genetic analysis; gene mapping;  
KW detection; homology; human; adrenal tissue; ds.

DT 15-FEB-1999 (first entry)  
XX EST clone BY66.  
XX  
XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
KW tissue growth; activin; inhibitor; chemotaxis; chemokinesis; haemostatic;  
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
XX gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
XX  
XX WO9845436-A2.  
XX  
XX 15-OCT-1998.  
XX  
XX 10-APR-1998; 98WO-US06955.  
XX  
XX 10-APR-1997; 97US-0838821.  
XX  
XX (GEM) GENETICS INST INC.  
XX  
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
PI Racie LA, Spaulding V, Treacy M;  
XX  
XX WPI; 1999-070077/06.  
XX  
XX New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries.  
XX  
XX Claim 1; Page 164; 618pp; English.  
XX  
XX The present sequence represents a human expressed sequence tag (EST).  
CC The polynucleotide, which is a secreted EST, and the encoded protein  
CC are predicted to have useful biological activities which would make  
CC them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals, although no supporting data is  
CC given. Suggested activities include nutritional activity, immune  
CC stimulating or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
XX  
XX Sequence 193 BP; 35 A; 39 C; 68 G; 51 T; 0 other;  
SQ

Query Match 32.6%; Score 84; DB 20; Length 193;  
Best Local Similarity 100.0%; Pred. No. 4.2e-14;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 175 CTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTC 234  
DB 193 CTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTTC 134  
QY 235 CTCGAGCAACACCCGCCCTTC 258  
DB 133 CTCGAGCAACACCCGCCCTTC 110  
RESULT 4  
QY7534  
ID Q77534 standard; DNA; 297 BP.  
XX  
XX Q77534;  
XX  
XX 23-SEP-1994 (first entry)  
XX  
XX Human genome fragment.  
XX  
XX Brain; placenta; bone marrow; genetic analysis; gene mapping;  
KW detection; homology; human; adrenal tissue; ds.

XX Homo sapiens.  
OS  
XX WO9401548-A.  
PN  
XX 20-JAN-1994.  
PD  
XX 13-JUL-1993; 93WO-GB01467.  
PF  
XX 13-JUL-1992; 92GB-0014857.  
PR  
XX (MEDI-) MEDICAL RES COUNCIL.  
PA  
XX Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;  
PI Sibson DR, Starkey M;  
PI WPI; 1994-035056/04.  
XX  
XX New nucleic acid fragment encoding gene products - can be used  
PT for genetic analysis and mapping  
PT  
XX Claim 1; Page 575-576; 616pp; English.  
XX  
XX Human nucleic acid fragments, isolated from brain, adrenal tissue,  
CC the placenta or bone marrow comprise any of: (A) a sequence  
CC selected from (Q76401-Q77613), (B) an allelic variation of a  
CC sequence as described in (A), or (C) a sequence complementary  
CC to (A) or (B).  
CC Preferred sequences exhibit no more than 90% homology to a human  
CC sequence known per se.  
XX  
XX Sequence 297 BP; 66 A; 117 C; 61 G; 52 T; 1 other;  
SQ

Query Match 17.1%; Score 44; DB 15; Length 297;  
Best Local Similarity 77.9%; Pred. No. 0.0026;  
Matches 53; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
QY 2 TCCGCGCCAGCTGCCAGAACCCGACGAGAACCCGACGAGGCTTCTTCTCTCTCTGTC 61  
DB 225 TCCGCGCCAGCTGCCAGAACCCGACGAGAACCCGACGAGGCTTCTTCTCTCTCTGTC 61  
QY 62 TCACCATC 69  
DB 285 TCGCCATC 292

RESULT 5  
V69699  
ID V69699 standard; DNA; 6354 BP.  
XX  
XX V69699;  
AC  
XX 08-FEB-1999 (first entry)  
DT  
XX 5' UTR sequence of iENTP DNA.  
DE  
XX  
XX Equilibrative nucleoside transport protein; iENTP; NBMPR; transport;  
KW nitrobenzylmercaptopurine riboside; antiviral; antitumour; screening;  
KW inhibitor; drug; knockout mice; cancer; chemotherapy; gene therapy;  
KW adenosine deaminase; factor VIII; UTR; untranslated region; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO9846749-A1.  
XX  
XX 22-OCT-1998.  
PD  
XX 10-APR-1998; 98WO-US07283.  
PF  
XX 09-APR-1998; 98US-0058389.  
PR  
XX 11-APR-1997; 97US-0838845.  
XX



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 21:29:29 ; Search time 77.57 Seconds  
(without alignments)  
1249.465 Million cell updates/sec

Title: US-08-799-910-9\_COPY\_211\_468

Perfect score: 258

Sequence: 1 GTCCGGCCCGCCAGTCCAGT.....AGCAACACCCGCCGCCATC 258.

Scoring table: IDENTITY\_NUC %  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: N\_Geneseq\_36:\*\*

1: /cgn2\_2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /cgn2\_2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /cgn2\_2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /cgn2\_2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /cgn2\_2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /cgn2\_2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /cgn2\_2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /cgn2\_2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /cgn2\_2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /cgn2\_2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /cgn2\_2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /cgn2\_2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /cgn2\_2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /cgn2\_2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /cgn2\_2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /cgn2\_2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /cgn2\_2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /cgn2\_2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /cgn2\_2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /cgn2\_2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /cgn2\_2/gcgdata/geneseq/geneseq/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query %	Score	Match	Length	DB ID	Description
1	258	100.0	1228	18	T94471	Human Fchd605 gene
2	258	100.0	1228	21	250711	Nucleotide sequenc
3	84	32.6	193	20	V89271	EST clone BY66. H
4	44	17.1	297	15	Q77534	Human genome fragm
5	38.6	15.0	6354	19	V69699	5' UTR sequence of
6	38.2	14.8	4228	21	Z51683	Human cyclic nucle
7	37.6	14.6	170	19	V69706	Nucleotide sequenc
8	36.4	14.1	1526	20	Z42235	Human normal bladd
9	36.2	14.0	1000	21	A02484	Human colon cancer
10	34.8	13.5	2953	19	V62462	Rat transcription
11	34.2	13.3	3690	13	Q30849	Type III procollag
12	33.8	13.1	4539	21	A08372	Human L-type amino

C 13	33.6	13.0	1120	15	Q70728	TATA-binding prote
C 14	33.6	13.0	1120	17	T42214	Drosophila TATA-bi
C 15	33.6	13.0	1120	18	T79601	TATA-binding prote
C 16	33.6	13.0	3765	13	Q31890	PMR-1 gene for det
C 17	33.6	13.0	5460	17	T16508	Vector pAC3A1 cont
C 18	33	12.8	1072	13	Q24285	Epo:IL-3 Flex, rec
C 19	33	12.8	11722	19	V34455	Human MHC class I
C 20	33	12.8	114955	20	X53491	Human adenosine A1
C 21	32.8	12.7	513	20	X91400	Human MHC class I
C 22	32.8	12.7	514	20	X91400	Human MHC class I
C 23	32.8	12.7	1450	21	Z47131	T. gondii immunoge
C 24	32.8	12.7	8169	19	V26609	Pseudomonas aerugi
C 25	32.4	12.6	201	8	N70194	Actinomadura hibis
C 26	32.4	12.6	201	8	N70195	Streptomyces prote
C 27	32.4	12.6	2681	20	X58429	Thermophilus therm
C 28	32.2	12.5	1524	21	A08393	Human L-type amino
C 29	32	12.4	742	21	Z45675	cDNA sequence of a
C 30	32	12.4	1459	21	A02528	Human colon cancer
C 31	32	12.4	3410	19	V61201	Full length cDNA s
C 32	32	12.4	3410	21	A06349	Prostate tumour sp
C 33	32	12.4	3410	21	A06349	Human immunogenic
C 34	32	12.4	34094	20	Z30163	Complete nucleotid
C 35	31.8	12.3	599	19	V43880	Mycobacterial meth
C 36	31.8	12.3	1321	21	A37128	Human PRO1781 (UNQ
C 37	31.6	12.2	1586	19	V34287	Human secreted pro
C 38	31.6	12.2	1907	19	V34314	Human cyclin-depen
C 39	31.6	12.2	2074	19	V32932	Human cyclin-depen
C 40	31.6	12.2	2074	19	V32933	Human cyclin-depen
C 41	31.4	12.2	595	14	Q50041	ICP34.5 fragment.
C 42	31.4	12.2	1335	13	Q23295	HSV-1 (F) ICP34.5
C 43	31.4	12.2	13842	21	Z87297	S. venezuelae macr
C 44	31.4	12.2	36778	21	Z87318	S. venezuelae pik
C 45	31.4	12.2	37948	21	Z87285	S. venezuelae pik

## ALIGNMENTS

RESULT 1  
ID T94471 standard; cDNA; 1228 BP.  
XX  
AC T94471;  
XX  
DT 03-MAR-1998 (first entry)  
XX  
DE Human Fchd605 gene differentially regulated in monocytes.  
XX  
KW Fchd602 gene; differential expression; monocyte; human;  
KW foam cell; cardiovascular disease; atherosclerosis; pschaemia;  
KW reperfusion; hypertension; restenosis; arterial inflammation;  
KW therapy; diagnosis; drug screening; marker; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..471  
ET /tag= a  
XX  
PN WO9730065-A1.  
XX  
PD 21-AUG-1997.  
XX  
PF 14-FEB-1997; 97WO-US02291.  
XX  
PR 13-FEB-1997; 97US-0799910.  
PR 16-FEB-1996; 96US-0011787.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Falb DA;  
XX  
DR WPI; 1997-424966/39.

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 10:53:09 ; Search time 1450.12 Seconds  
(without alignments)  
5934.126 Million cell updates/sec

Title: US-08-799-910-9  
Perfect score: 1228  
Sequence: 1 ATGTGTCACCTCTCGCAGCTG.....AAAAAAAAAAACTCGAG 1228

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: gb\_estl1:  
2: gb\_est2:  
3: gb\_est3:  
4: gb\_est4:  
5: gb\_est5:  
6: gb\_est6:  
7: gb\_est7:  
8: gb\_est8:  
9: gb\_est9:  
10: gb\_est10:  
11: gb\_est11:  
12: gb\_est12:  
13: gb\_est13:  
14: gb\_est14:  
15: gb\_est15:  
16: gb\_est16:  
17: gb\_est17:  
18: gb\_est18:  
19: gb\_est19:  
20: gb\_est20:  
21: gb\_est21:  
22: gb\_est22:  
23: gb\_est23:  
24: gb\_est24:  
25: gb\_est25:  
26: gb\_est26:  
27: gb\_est27:  
28: gb\_est28:  
29: gb\_est29:  
30: gb\_est30:  
31: gb\_est31:  
32: gb\_est32:  
33: gb\_est33:  
34: gb\_est34:  
35: gb\_est35:  
36: gb\_est36:  
37: gb\_est37:  
38: gb\_est38:  
39: gb\_est39:  
40: gb\_est40:  
41: em\_estba:  
42: em\_estfun:  
43: em\_esthum1:  
44: em\_esthum2:  
45: em\_esthum3:  
46: em\_esthum4:  
47: em\_esthum5:  
48: em\_esthum6:  
49: em\_esthum7:  
50: em\_esthum8:  
51: em\_esthum9:  
52: em\_esthum10:  
53: em\_esthum11:  
54: em\_esthum12:  
55: em\_esthum13:  
56: em\_esthum14:  
57: em\_esthum15:  
58: em\_esthum16:  
59: em\_esthum17:  
60: em\_esthum18:  
61: em\_esthum19:  
62: em\_esthum20:  
63: em\_estin1:  
64: em\_estin2:  
65: em\_estin3:  
66: em\_estin4:  
67: em\_estov1:  
68: em\_estov2:  
69: em\_estpl1:  
70: em\_estpl2:  
71: em\_estpl3:  
72: em\_estpl4:  
73: em\_estpl5:  
74: em\_estro1:  
75: em\_estro2:  
76: em\_estro3:  
77: em\_estro4:  
78: em\_estro5:  
79: em\_estro6:  
80: em\_estro7:  
81: em\_estro8:  
82: em\_estro9:  
83: em\_estro10:  
84: em\_estro11:  
85: em\_estro12:  
86: em\_estro13:  
87: gb\_est41:  
88: gb\_est42:  
89: gb\_est43:  
90: gb\_est44:  
91: gb\_est45:  
92: gb\_est46:  
93: gb\_est47:  
94: gb\_est48:  
95: gb\_est49:  
96: gb\_est50:  
97: gb\_est51:  
98: gb\_est52:  
99: gb\_est53:  
100: gb\_est54:  
101: gb\_est55:  
102: gb\_est56:  
103: gb\_est57:  
104: gb\_est67:  
105: gb\_est68:  
106: gb\_est69:  
107: gb\_est70:  
108: gb\_est71:  
109: gb\_est72:  
110: gb\_est73:  
111: gb\_est74:  
112: em\_esthum21:  
113: em\_esthum22:  
114: em\_esthum23:  
115: em\_estom1:  
116: em\_estom2:

117: em\_estpl6:\*  
118: em\_estpl7:\*  
119: em\_estpl8:\*  
120: em\_estpl9:\*  
121: em\_estpl10:\*  
122: em\_estpl11:\*  
123: em\_estpl12:\*  
124: em\_estpl13:\*  
125: em\_estpl14:\*  
126: em\_estpl15:\*  
127: em\_estpl16:\*  
128: em\_estpl17:\*  
129: em\_estpl18:\*  
130: em\_estpl19:\*  
131: em\_estpl20:\*  
132: em\_estpl21:\*  
133: em\_estpl22:\*  
134: em\_estpl23:\*  
135: em\_estpl24:\*  
136: em\_estpl25:\*  
137: em\_estpl26:\*  
138: em\_estpl27:\*  
139: em\_estpl28:\*  
140: em\_estpl29:\*  
141: em\_estpl30:\*  
142: em\_estpl31:\*  
143: em\_estpl32:\*  
144: em\_estpl33:\*  
145: em\_estpl34:\*  
146: em\_estpl35:\*  
147: em\_estpl36:\*  
148: em\_estpl37:\*  
149: em\_estpl38:\*  
150: em\_estpl39:\*  
151: em\_estpl40:\*  
152: em\_estpl41:\*  
153: em\_estpl42:\*  
154: em\_estpl43:\*  
155: em\_estpl44:\*  
156: em\_estpl45:\*  
157: em\_estpl46:\*  
158: em\_estpl47:\*  
159: em\_estpl48:\*  
160: em\_estpl49:\*  
161: em\_estpl50:\*  
162: em\_estpl51:\*  
163: em\_estpl52:\*  
164: em\_estpl53:\*  
165: em\_estpl54:\*  
166: em\_estpl55:\*  
167: em\_estpl56:\*  
168: em\_estpl57:\*  
169: em\_estpl58:\*  
170: em\_estpl59:\*  
171: em\_estpl60:\*  
172: em\_estpl61:\*  
173: em\_estpl62:\*  
174: em\_estpl63:\*  
175: em\_estpl64:\*  
176: em\_estpl65:\*  
177: em\_estpl66:\*  
178: em\_estpl67:\*  
179: em\_estpl68:\*  
180: em\_estpl69:\*  
181: em\_estpl70:\*  
182: em\_estpl71:\*  
183: em\_estpl72:\*  
184: em\_estpl73:\*  
185: em\_estpl74:\*  
186: em\_estpl75:\*  
187: em\_estpl76:\*  
188: em\_estpl77:\*  
189: em\_estpl78:\*

190: gb\_gss25:\*  
191: gb\_gss26:\*  
192: gb\_gss27:\*  
193: gb\_gss28:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	748.2	60.9	951	107	BE383865	BE383865 601273192
2	621.4	50.6	765	106	BE275966	BE275966 601120727
3	604.2	49.2	762	110	BE621346	BE621346 601493736
4	590.4	48.0	916	137	BE882130	BE882130 601505327
5	589.4	48.0	603	106	BE273386	BE273386 601143493
6	588.4	47.9	810	109	BE563592	BE563592 601334706
7	585.4	47.6	585	137	BE910077	BE910077 601502143
8	581.2	47.3	586	89	AW340002	AW340002 hc91d03.x
9	579.6	47.2	598	93	AW662348	AW662348 hi25c07.x
10	567.2	46.2	614	110	BE621600	BE621600 601493736
11	561.6	45.7	588	97	AW956285	AW956285 EST368355
12	557.4	45.4	617	38	AW690660	AW690660 AV690660
13	544.8	44.4	548	88	AW623252	AW623252 xn79g02.x
14	540.8	44.0	1070	137	BE878459	BE878459 601487866
15	519.4	42.3	521	25	AI800828	AI800828 wgl1f12.x
16	517.4	42.1	544	107	BE384949	BE384949 601276496
17	515.6	42.0	523	18	AI302813	AI302813 qn58g09.x
18	514.8	41.9	585	97	AW960290	AW960290 EST372361
19	513.4	41.8	518	12	AA806234	AA806234 oe29h02.s
20	512.4	41.7	513	40	AW150706	AW150706 xg38a10.x
21	506.2	41.2	524	22	AI566501	AI566501 tr60f07.x
22	505.4	41.0	509	9	AA582942	AA582942 nn80c01.s
23	503.4	41.0	521	23	AI685453	AI685453 tt88d12.x
24	501.6	40.8	510	89	AW337920	AW337920 xw63a07.x
25	496.4	40.4	553	142	N32077	N32077 yw97d08.s1
26	494.8	40.3	1155	135	BE786915	BE786915 601477783
27	487.6	39.7	548	96	AW887044	AW887044 RC1-OT008
28	486.8	39.6	535	25	AI802925	AI802925 tj58a07.x
29	486.6	39.6	494	15	AI022951	AI022951 ow66g09.s
30	482.4	39.3	484	19	AI336188	AI336188 qt44c12.x
31	481.4	39.2	491	39	AW081284	AW081284 xc30c01.x
32	481.4	39.2	492	10	AA643850	AA643850 np26f08.s
33	478.4	39.0	480	96	AW874420	AW874420 hq04a02.x
34	472.4	38.5	496	16	AI160053	AI160053 qc08a02.x
35	472.4	38.4	475	8	AA512938	AA512938 nh91b09.s
36	470.8	38.3	487	23	AI687081	AI687081 tp92h08.x
37	467.8	38.0	701	135	BE735733	BE735733 601304556
38	464.6	37.8	483	16	AI091958	AI091958 qa59d10.s
39	461.4	37.6	466	39	AW051587	AW051587 w97h05.x
40	458.4	37.3	460	17	AI185199	AI185199 qe35h09.s
41	458.4	37.3	460	24	AI738521	AI738521 w133a06.x
42	458.4	37.3	465	13	AA84985	AA84985 am35e07.s
43	453.6	36.9	461	12	AA812286	AA812286 nr82f07.s
44	451.6	36.8	459	147	W60982	W60982 zc98h09.s1
45	450.0	36.6	469	24	AI761431	AI761431 w965e08.x

## ALIGNMENTS

RESULT 1  
LOCUS BE383865 951 bp mRNA EST 21-JUL-2000  
DEFINITION 601273192F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:3614255 5', mRNA sequence.  
ACCESSION BE383865  
VERSION BE383865.1 GI:9329230  
KEYWORDS EST,  
SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 951)  
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: ATCC/DCTD/DTP  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCM276 row: g column: 24  
 High quality sequence stop: 747.  
 Location/Qualifiers  
 1..951  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3614255"  
 /clone\_lib="NIH\_MGC\_20"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dr priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(C). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 210 a 324 c 253 g 164 t

BASE COUNT  
 ORIGIN

Query Match 60.9%; Score 748.2; DB 107; Length 951;  
 Best Local Similarity 91.5%; Pred. No. 4.8e-180;  
 Matches 848; Conservative 0; Mismatches 73; Indels 6; Gaps 5;

QY 2 TGTGTCACTTCGACGTCGACCCGACCATGACCATCTCTGCGAGCCCGCCACCCCGGCC 61  
 DB 1 TGTGTCACTTCGACGTCGACCCGACCATGACCATCTCTGCGAGCCCGCCACCCCGGCC 60  
 QY 62 CTTCCACCATCCCGGACCCCGGGGGTCTCGGTCTCTGAGATCTTCACTTCGACCCCTC 121  
 DB 61 CTTCCACCATCCCGGACCCCGGGGGTCTCGGTCTCTGAGATCTTCACTTCGACCCCTC 120  
 QY 122 TCCCGGAGCCCGCAGCGGCCCTCGCGGGCCCGCCAGCGCTCTCGCGGCGCACCGAAAGC 181  
 DB 121 TCCCGGAGCCCGCAGCGGCCCTCGCGGGCCCGCCAGCGCTCTCGCGGCGCACCGAAAGC 180  
 QY 182 GCAGCCGACGAGGTTCTTACCTCTGAGTGTCTCGCGGCCAGCTGCCAGTCGAGGAACCGA 241  
 DB 181 GCAGCCGACGAGGTTCTTACCTCTGAGTGTCTCGCGGCCAGCTGCCAGTCGAGGAACCGA 240  
 QY 242 ACCGAGCCAAAGGCTCTCTTCTGCTGCTCACCATCGTCTTCTGCCAGATCCTGATGG 301  
 DB 241 ACCGAGCCAAAGGCTCTCTTCTGCTGCTCACCATCGTCTTCTGCCAGATCCTGATGG 300  
 QY 302 CTGAAGAGGCTGTGCGGGCCCTCTGCTCTCAGAGGAGCGCCCTAACGCGCGCATCCTCTGG 361  
 DB 301 CTGAAGAGGCTGTGCGGGCCCTCTGCTCTCAGAGGAGCGCCCTAACGCGCGCATCCTCTGG 360  
 QY 362 CGGCCACCCCTGTGTCTCCCGCTCTCGAGCCCTTTAATCTCACTTCGAGCCCTCGGACT 421  
 DB 361 CGGCCACCCCTGTGTCTCCCGCTCTCGAGCCCTTTAATCTCACTTCGAGCCCTCGGACT 420  
 QY 422 ACGCTCTGGACCTCAGGACTTTCTCTCAGCAACACCCGCGGCTTCTTAAGTGTGACTCC 481  
 DB 421 ACGCTCTGGACCTCAGGACTTTCTCTCAGCAACACCCGCGGCTTCTTAAGTGTGACTCC 480

QY 482 CGGACTCCCAAAAGAAATCCGAAAACACAAAGAAACACAGCGGTACCTGGTGGCC 541  
 DB 481 CGGACTCCCAAAAGAAATCCGAAAACACAAAGAAACACAGCGGTACCTGGTGGCC 540  
 QY 542 GAGAGCGTATCCCAACTCGGACTTCCGAGGCAACTTGAATCAGAACACTACAGCGGAG 601  
 DB 541 GAGAGCGTATCCCAACTCGGACTTCCGAGGCAACTTGAATCAGAACACTACAGCGGAG 599  
 QY 602 ACGCACCCCGTGTCTGAGCGGGGACCGAGGCGCAGAGAGCGGAGCGCATAGACCG 661  
 DB 600 ACGCACCCCGTGTCTGAGCGGGGACCGAGGCGCAGAGAGCGGAGCGCATAGACCG 659  
 QY 662 AGGCACAGCCAGC-TGGGGCTAGGCCCGGTGGGAAGAGAGAGCGTCTTAAATTAATTCT 720  
 DB 660 AGGCACAGCCAGCTTGGGGCTAGGCCCGGTGGGAAGAGAGAGCGTCTTAAATTAATTCT 719  
 QY 721 TATTGCTCTCTTAATTAATTAATTAATG--TATTATATAGTCTCTCTAGGTGATGAGAT 778  
 DB 720 TATTGCTCTCTTAATTAATTAATTAATG--TATTATATAGTCTCTCTAGGTGATGAGAT 779  
 QY 779 GTGAGCTAATTAATTAATTAATTAATTAATGAGGCTGTGAGATGTTCCCTCTCTGTAAT 838  
 DB 780 GTGAGCTAATTAATTAATTAATTAATTAATTAATGAGGCTGTGAGATGTTCCCTCTCTGTAAT 837  
 QY 839 GCAGGCTCTCTGTTGTTTATTTAGCTTTTGTGGACTTGTGGAGAGGAGGACCTGGAACT 898  
 DB 838 AGGCACAGGATTAATTAATTAATTAATTAATTAATTAATGAGGCTGTGGAGAGGAGGAACT 897  
 QY 899 GCGGCAAGTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 925  
 DB 898 AGGAGCAACGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 924

RESULT 2  
 BE275966 765 bp mRNA 13-JUL-2000  
 LOCUS 601120727F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:2967534 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE275966  
 VERSION BE275966.1 GI:9150928  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 765)  
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: ATCC/DCTD/DTP  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCM72 row: e column: 07  
 High quality sequence stop: 7  
 High quality sequence start: 699.  
 Location/Qualifiers  
 1..765  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2967534"  
 /clone\_lib="NIH\_MGC\_20"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dr priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'



```
/note="Organ: kidney"
```



cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 112 a 241 c 154 g 96 t

ORIGIN

Query Match 48.08; Score 589.4; DB 106; Length 603;  
Best Local Similarity 99.1%; Pred. No. 1.3e-139;  
Matches 601; Conservative 0; Mismatches 1; Indels 1; Caps 1;

QY 35 CCATCTCTGACGGCCGACCCGCGCCCTCCACCATCCGCGGACCCGCGGGGCTCCG 94  
Db 1 CCATCTCTGACGGCCGACCCGCGCCCTCCACCATCCGCGGACCCGCGGGGCTCCG 60

QY 95 GTCTTGAGATCTTACCTTTCGACCTCTCCCGAGCCGCGGCGCCCTGCGGCGCC 154  
Db 61 GTCTTGAGATCTTACCTTTCGACCTCTCCCGAGCCGCGGCGCCCTGCGGCGCC 120

QY 155 CCAGCGCTCTCCGCGGACCGAAGCGGACCGGAGGTTCTTACCTTCGAGTGTCC 214  
Db 121 CCAGCGCTCTCCGCGGACCGAAGCGGACCGGAGGTTCTTCTTCTGCTCTCA 274

QY 215 GCGCGCAGTTCGAGGACCGAAGCGGACCGGAGGTTCTTCTTCTGCTCTCA 274  
Db 181 GCGCGCAGTTCGAGGACCGAAGCGGACCGGAGGTTCTTCTTCTGCTCTCA 240

QY 275 CCATGCTCTTCTGCCAGATCTGATGGCTGAAGAGGTTCTGCGGCGCCCTGCCCTCCAG 334  
Db 241 CCATGCTCTTCTGCCAGATCTGATGGCTGAAGAGGTTCTGCGGCGCCCTGCCCTCCAG 300

QY 335 AGGACGCCCTAACGCGGACCTGCGGCGCCACCCCTGTGTCC--CCGCTCTCGAGCCC 393  
Db 301 AGGACGCCCTAACGCGGACCTGCGGCGCCACCCCTGTGTCTTCTGCGGCGCC 360

QY 394 TTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACTCTGACCTTCTTCTCCAGCAA 453  
Db 361 TTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACTCTGACCTTCTTCTCCAGCAA 420

QY 454 CACCGCGCGCCCTTCTAAGTCTGACTCTCCCGGACCTCCGCAAGAAATCCGAAACAC 513  
Db 421 CACCGCGCGCCCTTCTAAGTCTGACTCTCCCGGACCTCCGCAAGAAATCCGAAACAC 480

QY 514 AAGAAACACGAGGCTACCTGGTGGGAGAGGTTCTGCGGCGCCCTGCCCTCCAGCAA 573  
Db 481 AAGAAACACGAGGCTACCTGGTGGGAGAGGTTCTGCGGCGCCCTGCCCTCCAGCAA 540

QY 574 AACTTGAACTGAACTACAGCGGAGACCCGCGGTTGAGGCGGACCGAGGC 633  
Db 541 AACTTGAACTGAACTACAGCGGAGACCCGCGGTTGAGGCGGACCGAGGC 600

QY 634 CCA 636  
Db 601 GCA 603

RESULT 6  
BE563592 810 bp mRNA EST 15-AUG-2000  
LOCUS 601.14706F1 NIH\_MGC\_39 Homo sapiens cDNA clone IMAGE:3688825 5',  
DEFINITION mRNA sequence.  
ACCESSION BE563592  
VERSION BE563592.1 GI:9807312  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 810)  
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

# TITLE JOURNAL COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCN382 row: k column: 02  
High quality sequence stop: 692.

## FEATURES source

1..810  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3688825"  
/clone\_lib="NIH\_MGC\_39"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 176 a 284 c 205 g 145 t  
ORIGIN

Query Match 47.9%; Score 588; DB 109; Length 810;  
Best Local Similarity 92.1%; Pred. No. 3e-139;  
Matches 753; Conservative 0; Mismatches 50; Indels 15; Gaps 12;

QY 1 ATGTGTCTACTTCGAGCTGCGGAGCCGCGGCGGCTTCCGGTCTTGAGATCTTCACCTTCGACCT 120  
Db 1 ATGTGTCTACTTCGCGAGTGC--ACCCGACATGACCATCTCTGCGGCGCCGACCCCGG-C 57

QY 61 CCCCTCCACCATCCCGGAGCCCGCGGCGGCTTCCGGTCTTGAGATCTTCACCTTCGACCT 120  
Db 58 CCCCTCCACCATCCCGGAG--CCCGGCGGCGCTTCCGGTCTTGAGATCTTCACCTTCGACCT 116

QY 121 CTCCCGGAGCCCGGAGCGCCCTCTGCGGCGCCCGCTGCGGCGCCCGCTGCGGCGCGGCGGCGG 180  
Db 117 CTCCCGGAGCCCGGAGCGCCCTCTGCGGCGCCCGCTGCGGCGCCCGCTGCGGCGCGGCGG 176

QY 181 CGCAGCGCAGGGTTCTTACCTCTGAGTGTGCGGCGCCAGCTGCCAGTCGAGGAGCG 240  
Db 177 CGCAGCGCAGGGTTCTTACCTCTGAGTGTGCGGCGCCAGCTGCCAGTCGAGGAGCG 236

QY 241 AACCCAGCCAAAGGCTTCTTCTTCTGCTCACCATCTGCTTCTGCGGAGTCCGTGATG 300  
Db 237 AACCCAGCCAAAGGCTTCTTCTTCTGCTCACCATCTGCTTCTGCGGAGTCCGTGATG 296

QY 301 GCTGAAGAGGTTGTGCGGCGCCCTGCTCCAGAGGAGCGCCCTAACGCGGCAATCCC-T 359  
Db 297 GCTGAAGAGGTTGTGCGGCGCCCTGCTCCAGAGGAGCGCCCTAACGCGGCAATCCCCT 356

QY 360 GCGCGCCACCCCTGTGTGCGGCGCTTCTGAGGCGCTTAAATCTGACTTCGAGGCGCTCGGA 419  
Db 357 GCGCGCCACCCCTGTGTGCGGCGCTTCTGAGGCGCTTAAATCTGACTTCGAGGCGCTCGGA 416

QY 420 CTACGCTCTGAGCTCAGCACTTTCCTCCAGCAACACCCCGCGCTTCTTAACCTGTGACT 479  
Db 417 CTAGGCTCTGAGCTCAGCACTTTCCTCCAGCAACACCCCGGAC--GCTTCTTAACCTGTGACT 474

QY 480 CCCCAGCACTCCCCAAAGAAATCCGAAACACCAAGAAACACCAAGAAACACCAAGGCGTACCTGTC 539  
Db 475 CCCCAGCACTCCCCAAAGAAATCCGAAACACCAAGAAACACCAAGGCGTACCTGTCGTC 534

```

QY 540 GCAGAGCGTATCCCACTGG-GACTTCGAGGCAACTTGAACCTCAGAAC-ACTACAGC 597
DB 535 GCAGAGCGTATCCCACTGGAGACTTCGAGGCAACTTGAACCTCAGAACACTACAGC 594
QY 598 GGAGAGCGCCACCC--GGTGCTTCAGGCGGACCGAGCGCA-CAGAGACTGAGG-CGCAAT 653
DB 595 GGAGAGCGCCAACTGGTGCTGTGAGGCGGACCGAGCGCA-CAGAGACTGAGGCGCAT 654
QY 654 AGAGACCGAGGACACGCCAGCTGGGCTAGGCCCGGTGGGAAGGAGAGCGCTGTTAAAT 713
DB 655 AGAGACCGAGGACAG-CCAGATGGGCTAGGCCCGGTGGGAAGGAGAGCGTCCGCAAT 713
QY 714 TATTTCTATTGCTCTAATAATATATATATATATATATATATATATATATATATATAT 773
DB 714 TAATATCTTATGCTTCCTAACTAACTAATAATATGTA-TTATGTACGCTCCCTAGGTGATG 772
QY 774 GAGATGCTACGTAATATATTTTAACTTATCAAGG 811
DB 773 GACAAATTGTACGTAATATATTTAACTATGCCAGG 810

RESULT 7
BE910077 BE910077 585 bp mRNA EST 29-SEP-2000
LOCUS 601502143F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3903754 5',
DEFINITION mRNA sequence.
ACCESSION BE910077
VERSION BE910077.1 GI:10406309
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 585)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM695 row: f column: 11
High quality sequence stop: 585.
FEATURES
Location/Qualifiers
1..585
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3903754"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 131 a 119 c 188 g 147 t
ORIGIN

```

```

QY 604 GCCACCCGCTGCTTGAGGCGGACCGACGACAGACCGGCGCATAGAGACCGAG 563
DB 61 GCCACCCGCTGCTTGAGGCGGACCGACGACAGACCGGCGCATAGAGACCGAG 120
QY 664 GCACAGCCAGCTGGGCGCTAGGCCCGGTGGGAAGGAGAGCGCTGTTAAATTTATTTCTTAT 723
DB 121 GCACAGCCAGCTGGGCGCTAGGCCCGGTGGGAAGGAGAGCGCTGTTAAATTTATTTCTTAT 180
QY 724 TGCCTCTAAATTAATATTTATATATATATATATATATATATATATATATATATATAT 783
DB 181 TGCCTCTAAATTAATATTTATATATATATATATATATATATATATATATATATATAT 240
QY 784 CGTAATATATTTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTGCTGTAATATGAGG 843
DB 241 CGTAATATTTATTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTGCTGTAATATGAGG 300
QY 844 TCTCTTGGTATTTATTTAGCTTTGTGGAGCTTGTGGAAGCAGGACACCTGGAACCTCGGC 903
DB 301 TCTCTTGGTATTTATTTAGCTTTGTGGAGCTTGTGGAAGCAGGACACCTGGAACCTCGGC 360
QY 904 AAAGTAGGAGAGAAATGGGAGGACTCGGCGTGGGAGGACGCTCCCGGCTGGGATCAAG 963
DB 361 AAAGTAGGAGAGAAATGGGAGGACTCGGCGTGGGAGGACGCTCCCGGCTGGGATCAAG 420
QY 964 TCTGGTGGTGGTGGTAAAGTTTAGGAGGTGACTGCAATCTCCAGCATCTCAACTCCGCT 1023
DB 421 TCTGGTGGTGGTGGTAAAGTTTAGGAGGTGACTGCAATCTCCAGCATCTCAACTCCGCT 480
QY 1024 GTCTACTGTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCTTCCATCT 1083
DB 481 GTCTACTGTGTGAGACTTCGGCGGACCATTAGGAATGAGATCCGTGAGATCCTTCCATCT 540
QY 1084 TCTTTGAAGTCGCTTTAGGCTGCTCGAGGTAGAGGTTGGGG 1128
DB 541 TCTTTGAAGTCGCTTTAGGCTGCTCGAGGTAGAGGTTGGGG 585

RESULT 8
AW340002/c
LOCUS AW340002 586 bp mRNA EST 31-JAN-2000
DEFINITION hc91d03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2907365 3', mRNA sequence.
ACCESSION AW340002
VERSION AW340002.1 GI:6836628
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 586)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CCAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Gibco
High quality sequence stop: 471.
FEATURES
Location/Qualifiers
1..586
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2907365"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p77T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NIH, and B-cell

```

NCI\_CGAP\_CCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 155 a 187 c 110 g 134 t

Query Match 47.3%; Score 581.2; DB 89; Length 586;  
Best Local Similarity 99.3%; Pred. No. 1.5e-137; Mismatches 0; Gaps 0;  
Matches 583; Conservative 0; Indels 0; Gaps 0;

Qy 621 GCGGACCGGACGACAGAGACGAGCGGCATAGACAGCCAGGCGCAGCCAGCTGGGG 680  
Db 586 GCGGACCGGACGACAGAGACGAGCGGCATAGACAGCCAGGCGCAGCCAGCTGGGG 527  
Qy 681 CTAGGCCGCGTGGGAAGGAGCGTCCTTAATTTATTTCTTATTTCTCTCTAATTAATTT 740  
Db 526 CTAGGCCGCGTGGGAAGGAGCGTCCTTAATTTATTTCTTATTTCTCTCTAATTAATTT 467  
Qy 741 TATATGATTTATCTAGCTCTCTCTAGCTGATGAGATGTTAGTAAATATTTATTTAA 800  
Db 466 TATATGATTTATCTAGCTCTCTCTAGCTGATGAGATGTTAGTAAATATTTATTTAA 407  
Qy 801 CTTATGCAAGGCTGAGATGTTCCCTCTCTGTAATGCAAGTCTCTTGGTATTTATTG 860  
Db 406 CTTATGCAAGGCTGAGATGTTCCCTCTCTGTAATGCAAGTCTCTTGGTATTTATTG 347  
Qy 861 AGCTTTGTGGACTGTGGGAAGCAGGACACCTTGAACCTGGCGCAAGTAGAGAGAANAAT 920  
Db 346 AGCTTTGTGGACTGTGGGAAGCAGGACACCTTGAACCTGGCGCAAGTAGAGAGAANAAT 287  
Qy 921 GCGGAGACTCGGCTGGGGGAGGACGTCCTGGGTGGGATGAAGTCTGGTGGTGGTGGT 980  
Db 286 GCGGAGACTCGGCTGGGGGAGGACGTCCTGGGTGGGATGAAGTCTGGTGGTGGTGGT 227  
Qy 981 AGTTAGAGGTGACTGCATCTCCAGCATCTCAACCTCCGCTCTCTTCTTGAAGTCCCTTTA 1040  
Db 226 AGTTAGAGGTGACTGCATCTCCAGCATCTCAACCTCCGCTCTCTTCTTGAAGTCCCTTTA 167  
Qy 1041 TCGGCGGACCATTAGGAATGAGATCCGTGAGATCTTCCATCTCTTCTTGAAGTCCCTTTA 1100  
Db 166 TCGGCGGACCATTAGGAATGAGATCCGTGAGATCTTCCATCTCTTCTTGAAGTCCCTTTA 107  
Qy 1101 GCGTGGCTGCGAGGTAGAGGCTGGGGGTTGGTGGGCTGTCCAGGACGACTGTGAGAT 1160  
Db 106 GCGTGGCTGCGAGGTAGAGGCTGGGGGTTGGTGGGCTGTCCAGGACGACTGTGAGAT 47  
Qy 1161 GCGCTAGTATGTTCTGTGAACACAAATAAATGATTACTGTGCA 1206  
Db 46 GCGCTAGTATGTTCTGTGAACACAAATAAATGATTACTGTGCA 1

RESULT 9  
AM662348/c 598 bp mRNA EST 06-APR-2000  
LOCUS h125c07.x1 NCI\_CGAP\_Co14 Homo sapiens cdna clone IMAGE:2973324 3',  
DEFINITION mRNA sequence.  
ACCESSION AM662348  
VERSION AM662348.1 GI:7454886  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 598)  
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITILE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL COMMENT  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCAP clone distribution Information can be found through the I.M.A.G.E. Consortium/LLNL at:  
image.llnl.gov/image/html/iresources.shtml  
Seq primer: -40UP from Gibco  
High quality sequence stop: 392.  
Location/Qualifiers  
1. .598  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2973324"  
/clone\_lib="NCI\_CGAP\_Co14"  
/issue\_type="moderately-differentiated adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.7 kb. Life Technologies catalog #:  
11531-019"

BASE COUNT 163 a 187 c 116 g 132 t  
ORIGIN  
Query Match 47.2%; Score 579.6; DB 93; Length 598;  
Best Local Similarity 98.5%; Pred. No. 3.9e-137; Mismatches 0; Indels 0; Gaps 0;  
Matches 585; Conservative 0; Indels 0; Gaps 0;

Qy 611 GGTGCTTGAGCGGGACCGAGCGCCACAGAGCCGAGCGCATAGAGCCGAGGACGACG 670  
Db 598 GGTGCTTGAGCGGGACCGAGCGCCACAGAGCCGAGCGCATAGAGCCGAGGACGACG 539  
Qy 671 CGAGCTGGGCTAGGCGCGGTGGGAAGGAGAGCGTGGTAAATTTATTTCTTATTGCTCT 730  
Db 538 CGAGCTGGGCTTAGGCGCGGTGGGAAGGAGAGCGTGGTAAATTTATTTCTTATTGCTCT 479  
Qy 731 AATTAATATTTATATGTTATTTATGTTACCTCTCTAGGTGATGAGATGTGTAAGTAA 790  
Db 478 AATTAATATTTATATGTTATTTATGCGAGTCTCTTAGGTGATGAGATGTGTAAGTAA 419  
Qy 791 TTTATTTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTGCTCTGCTCTGCTCTGCTCT 850  
Db 418 TTTATTTTAACTTATGCAAGGGTGTGAGATGTTCCCTCTGCTCTGCTCTGCTCTGCTCT 359  
Qy 851 GTATTTTATGAGCTTTGTGGAGTGGTGGGAAGCAGCACACCTGGAACCTGGGCAAGTAG 910  
Db 358 GTATTTTATGAGCTTTGTGGAGTGGTGGGAAGCAGCACACCTGGAACCTGGGCAAGTAG 299  
Qy 911 GAGAAGAAATGGGAGGACTCGGCTGGGGGAGGACCTCCGCGCTGGGATCAAGTCTGCTG 970  
Db 298 GAGAAGAAATGGGAGGACTCGGCTGGGGGAGGACCTCCGCGCTGGGATCAAGTCTGCTG 239  
Qy 971 GTGGGTGTAAGTTTAGGAGGTGACTGCAATCTCCAGCATCTCAACTCCGCTCTGCTACT 1030  
Db 238 GTGGGTGTAAGTTTAGGAGGTGACTGCAATCTCCAGCATCTCAACTCCGCTCTGCTACT 179  
Qy 1031 GTGTGAGACTTCGCGGACCATTTAGGAATGAGATCCGTGAGATCCCTTCCATCTCTTGAA 1090  
Db 178 GTGTGAGACTTCGCGGACCATTTAGGAATGAGATCCGTGAGATCCCTTCCATCTCTTGAA 119  
Qy 1091 GTCGCCTTTAGGTGGCTGCGAGGTAGAGGGTTGGGGGTTGGTGGGCTGTCCAGGACGA 1150  
Db 118 GTCGCCTTTAGGTGGCTGCGAGGTAGAGGGTTGGGGGTTGGTGGGCTGTCCAGGACGA 59  
Qy 1151 CTGTCGAGATCGCTAGTATGTTCTGTGAACACAAATAAATGATTACTGTCT 1204  
Db 58 CTGTCGAGATCGCTAGTATGTTCTGTGAACACAAATAAATGATTACTGTCT 5

```
RESULT 10
BE621600/c
LOCUS
DEFINITION
601493736T1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3895736 3',
mRNA sequence.
ACCESSION
BE621600
VERSION
BE621600.1 GI:9892540
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 614)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue procurement: ATCC
cDNA library preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUCM674 row: h column: 09
High quality sequence start: 12
High quality sequence stop: 613.
FEATURES
source
1..614
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3895736"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 158 a 198 c 123 g 135 t
ORIGIN
Query Match 46.2%; Score 567.2; DB 110; Length 614;
Best Local Similarity 98.2%; Pred. No. 5.7e-134;
Matches 595; Conservative 0; Mismatches 8; Indels 3; Gaps 2;
Qy 582 CTCAGACACTACAGCGGACGACGCCACCCGGTCTTTCAGCGGGACCCAGCGGCACACAG 641.
Db 614 CTCAGACACTACAGCGGACGACGCCACCCGGTCTTTCAGCGGGACCCAGCGGCACACAG 555
Qy 642 ACCGAGCGCATAGACAGCCAGGACAGCCAGCCAGCTGGGGGTAGGCCCGGTGGGAAGGAGA 701
Db 554 ACCGAGCGCATAGACAGCCAGGACAGCCAGCCAGCTGGGGGTAGGCCCGGTGGGAAGGAGA 495
Qy 702 GCGTCGTAA--TTTATTCTTATTGCTCTTAATAATATATATATATATATATATATATAT 759
Db 494 GCGTCGTAAATGTTATGTTCTTATGCTCTTAATAATATATATATATATATATATATAT 435
Qy 760 CTTCTAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 819
Db 434 CTTCTAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 375
Qy 820 TGTTCCTCTGCTGCTAAATCAGGCTCTCTTGGTATTTATTGAGCTTTTGTGGAGCTGGTGG 879
Db 374 TGTTCCTCTGCTGCTAAATCAGGCTCTCTTGGTATTTATTGAGCTTTTGTGGAGCTGGTGG 315
Qy 880 AAGCAGGACACCTCGACCTCGGGCAAGTAGGAGAGAAATGGGGAGA-CTCGGGGTGGG 938
|||||
```

```
Db 314 AAGCAGGACACCTGGAACTCGGCAAGTAGGAGAGAAATGGGAGGATCTCGGGTGGG 255
Qy 939 GGAGGACGTCCTCCGCTCGGATGAAGTCTGGTGGTGGTAACTTTAGGAGGTCACATGC 998
|||||
Db 254 GGAGGACGTCCTCCGCTCGGATGAAGTCTGGTGGTGGTAACTTTAGGAGGTCACATGC 195
Qy 999 ATCTCCAGGATCTCAACTCCGCTCTGCTACTGTGTGAGACTTCGGCGGACCATAGGAA 1058
|||||
Db 194 ATCTCCAGGATCTCAACTCCGCTCTGCTACTGTGTGAGACTTCGGCGGACCATAGGAA 135
Qy 1059 TGAGATCCGTGAGATCCCTTCATCTTCTTGAAGTCGCTTTAGGGTGGCTCGGAGGTAGA 1118
|||||
Db 134 TGAGATCCGTGAGATCCCTTCATCTTCTTGAAGTCGCTTTAGGGTGGCTCGGAGGTAGA 75
Qy 1119 GGGTGGGGTGGTGGGCTGTACAGGAGGACCTGTGAGATCCGCTAGTATCTTCTCTG 1178
|||||
Db 74 GGGTGGGGTGGTGGGCTGTACAGGAGGACCTGTGAGATCCGCTAGTATCTTCTCTG 15
Qy 1179 AACACA 1184
|||||
Db 14 CACAA 9
RESULT 11
AW956285
LOCUS
DEFINITION
EST368355 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.
ACCESSION
AW956285
VERSION
AW956285.1 GI:8145968
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 588)
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holl
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 90
Seq primer: Reverse.
Location/Qualifiers
1..588
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGD"
/note="Vector: pBluescriptSkm"
BASE COUNT 135 a 117 c 187 g 149 t
ORIGIN
Query Match 45.7%; Score 561.6; DB 97; Length 588;
Best Local Similarity 98.3%; Pred. No. 1.5e-132;
Matches 578; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
Qy 556 AACTGGGCACTTCCGAGGCAACTTGAACCTCAGAACACTACAGCGGAGACGCCACCGTGC 615
Db 1 AACTGGGCACTTCCGAGGCAACTTGAACCTCAGAACACTACAGCGGAGACGCCACCGTGC 60
Qy 616 TTGAGCGGGACCCAGGCGCACAGACCGAGCGGCATAGAGACCGAGGACACAGCCGAGC 675
Db 61 TTGAGCGGGACCCAGGCGCACAGACCGAGCGGCATAGAGACCGAGGACACAGCCGAGC 120
Qy 676 TGGGGCTAGGCCCGGTCGGGAAGGAGACGCTGTTTATTTCTTATTTCTCTAATTA 735
|||||
```

```
Db 121 TGGGCTAGCCGCGTGGGAAGGAGAGCGTCTTAATTTATTTCTTATTTGCTCCCTAATTA 180
QY 736 ATATTATATGATTTATGATGCTCTCTAGGTGATGAGATGCTAGTAAATATTTAT 795
Db 181 ATATTATATGATTTATGATGCTCTCTAGGTGATGAGATGCTAGTAAATATTTAT 240
QY 796 TTTAACTTATGCAAGGCTGTGAGATGTTCCCTCTCTGTAATGAGCTCTCTGCTAT 855
Db 241 TTTAACTTATGCAAGGCTGTGAGATGTTCCCTCTCTGTAATGAGCTCTCTGCTAT 300
QY 856 TATTGAGCTTTGTTGGAGCTGTGGAAGCAGGACACCTTGAACCTGCGGCAAAAGTAGGAA 915
Db 301 TATTGAGCTTTGTTGGAGCTGTGGAAGCAGGACACCTTGAACCTGCGGCAAAAGTAGGAA 360
QY 916 GAAATGGGAGGACTCGGCTGGGAGGAGGACGCTCCGGCTGGAGTAACTCTGGTGGTGG 975
Db 361 GAAATGGGAGGACTCGGCTGGGAGGAGGACGCTCCGGCTGGAGTAACTCTGGTGGTGG 420
QY 976 TCGTAAGTTTAGGAGCTGACTGCTATCCCTCCAGCATCTCAACTCCGCTCTGCTACTGCTG 1035
Db 421 TCGTAAGTTTAGGAGCTGACTGCTATCCCTCCAGCATCTCAACTCCGCTCTGCTACTGCTG 480
QY 1036 AGACTTCGGCGGACCATTAGGAATGAGATCCGCTGAGATCCCTTCCATCTCTTGAAGTCG 1095
Db 481 AGACTTCGGCGGACCATTAGGAATGAGATCCGCTGAGATCCCTTCCATCTCTTGAAGTCG 540
QY 1096 CTTTAGGCTGCTCGGAGGT-AGAGGGTTGGGGTGGTGGTGGTGGTGA 1142
Db 541 CTTTAGGCTGCTCGGAGGT-AGAGGGTTGGGGTGGTGGTGGTGGTGA 588

RESULT 12
AV690660 617 bp mRNA EST 25-SEP-2000
LOCUS AV690660 GK Homo sapiens cDNA clone GKCBA04 5', mRNA sequence.
DEFINITION AV690660
VERSION AV690660.1 GI:10292523
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Wu, T., Qian, B., Huang, Q., Kang, B., Gao, X., Xu, Z., Xiao, H.,
Xu, X., Li, N., Peng, Y., Liu, F., Ou, J., Song, H., Cheng, Z., Zeng, L.,
Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang
Y., Gu, Y., Chen, Z., and Han, Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@hgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1. 617
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCBA04"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/notes=Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
131 a 182 c 164 g 139 t 1 others
BASE COUNT
ORIGIN

Query Match 45.4%; Score 557; DB 38; Length 617;
```

Best Local Similarity 95.5%; Pred. No. 2.3e-131;  
Matches 596; Conservative 0; Mismatches 21; Indels 7; Gaps 2;

```
QY 277 ATCGTCTTCTCCAGATCCTGATGGCTGAAGAGGCTGTGCGGGCCCTTCCCTCCAGAG 336
Db 1 ATCGTCTTCTCCAGATCCTGATGGCTGAAGAGGCTGTGCGGGCCCTTCCCTCCAGAG 60
QY 337 GACGCCCTTAACGCGGATCCTTGGCGCCACCCCTGTGTCGCCCGTCCCTCGAGCCCTTT 396
Db 61 GACGCCCTTAACGCGGATCCTTGGCGCCACCCCTGTGTCGCCCGTCCCTCGAGCCCTTT 120
QY 397 AATCTGACTTTCGAGCCCTTCGGACTTACGCTCTGACCTTCAGACCTTTCCTCCAGCAAC 456
Db 121 AATCTGACTTTCGAGCCCTTCGGACTTACGCTCTGACCTTCAGACCTTTCCTTCAGCAAC 180
QY 457 CCGGCCCTTCTTAACCTGCTGACTCCCGCAGCTCCCGCAAAAGATCGCAAAACACAAA 516
Db 181 CCGGCCCTTCTTAACCTGCTGACTTCCCGCAGCTTCCCGCAAAAGATCGC---ACCACAAA 236
QY 517 GAAACACGAGCGTACCTTGGTTCGCCGAGAGCGTATCCCGCAACTCGGACTTCCGAGCGCAAC 576
Db 237 GAAACACGAGCGTACCTTGGTTCGCCGAGAGCGTATCCCGCAACTCGGACTTCCGAGCGCAAC 296
QY 577 TTGAATCAGAACACTACGCGGAGAGCGCCACCCGGTCTTTGAGCGCGGACCGAGCGCA 636
Db 297 TTGAATCAGAACACTACGCGGAGAGCGCCACCCGGTCTTTGAGCGCGGACCGAGCGCA, 356
QY 637 CAGAGACGAGGCGCATAGACGCGGACGCGGACGCGGCTAGGGCTAGGGCCGCTGGGAA 696
Db 357 CAGAGACGAGGCGCATAGACGCGGACGCGGACGCGGCTAGGGCTAGGGCCGCTGGGAA 416
QY 697 GGAGAGCGTCTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 756
Db 417 GGAGAGCGTCTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 476
QY 757 CGTCTCTAGGTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 816
Db 477 CGTCTCTAGGTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 536
QY 817 AGATGTTCCCTCTGTTAAATGAGTCTCTTGGTATTTATTTATTTATTTATTTATTTATTT 876
Db 537 AGATGTTCCCTCTGTTAAATGAGTCTCTTGGTATTTATTTATTTATTTATTTATTTATTT 1
QY 877 TGAAGCAGGACGACCTTGGAACTGC 900
Db 594 GGAAGCAGGACGACCTTGGCTGCGGC 617
```

RESULT 13

AW263252/c

LOCUS

DEFINITION

AW263252

AW263252

AW263252.1

GI:6640068

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 548)

NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from Gibco

High quality sequence stop: 466.

Location/Qualifiers

```
source          1. 548
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2700722"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pMT3D-Pac (pharmacia) with
a modified polylinker; Site_1: Not 1; Site_2: ECO RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell.
NCL CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT      157 a 171 c 98 g 122 t
ORIGIN
Query Match      44.4%; Score 544.8; DB 88; Length 548;
Best Local Similarity 99.6%; Pred. No. 2.8e-128;
Matches 546; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 660 CGAGGCACACCCAGCTGGGCGTAGGCCGCTGGGAGGAGAGCGTCGTTAATTATTC 719
DB 548 CGAGGCACACCCAGCTGGGCGTAGGCCGCTGGGAGGAGAGCGTCGTTAATTATTC 489
QY 720 TTATTCCTCTAATTAATATTTATATGATTTATGACGTCCTAGGTGAGATG 779
DB 488 TTATTCCTCTAATTAATATTTATGATTTATGACGTCCTAGGTGAGATG 429
QY 780 TGACGTAATATTTATTTAACTTATGCAAGGCTGAGATGTTCCCTCTGCTTAATG 839
DB 428 TGACGTAATATTTATTTAACTTATGCAAGGCTGAGATGTTCCCTCTGCTTAATG 369
QY 840 CAGGTCCTCTTGGTATTTATGAGCTTTGTGGGACTGTGTGAAGCAGACACCTGGAACTG 899
DB 368 CAGGTCCTCTTGGTATTTATGAGCTTTGTGGGACTGTGTGAAGCAGACACCTGGAACTG 309
QY 900 CGCAAAAGTAGAGACAAATAGGGAGGACTCGGGTGGGGGAGGAGCTCCCGGCTGGAT 959
DB 308 CGCAAAAGTAGAGACAAATAGGGAGGACTCGGGTGGGGGAGGAGCTCCCGGCTGGAT 249
QY 960 GAAGTCTGGTGGTGGTGAAGTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCC 1019
DB 248 GAAGTCTGGTGGTGGTGAAGTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCC 189
QY 1020 GTCTGTCTACTGTGTGAGACTTCGGCGGACATTAGGAATGAGATCCGTGAGATCTTCC 1079
DB 188 GTCTGTCTACTGTGTGAGACTTCGGCGGACATTAGGAATGAGATCCGTGAGATCTTCC 129
QY 1080 ATCTCTTGAAGTCGCCCTTTAGGGTGGCTCCGAGGTAGAGGGTTGGGGTGGGGCTG 1139
DB 128 ATCTCTTGAAGTCGCCCTTTAGGGTGGCTCCGAGGTAGAGGGTTGGGGTGGGGCTG 69
QY 1140 TCAGGAGGCACTGTCGAGATCGCTAGTATGTTCTGTGAACAAATAAATGATTTA 1199
DB 68 TCAGGAGGCACTGTCGAGATCGCTAGTATGTTCTGTGAACAAATAAATGATTTA 9
QY 1200 CTGTCAAA 1207
DB 8 CTGTCAAA 1
```

```
RESULT 14
BE878459
LOCUS
DEFINITION BE878459 1070 bp mRNA EST 27-SEP-2000
ACCSSION BE878459
```

```
VERSION BE878459.1 GI:10327235
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/DTB/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM660 row: c column: 17
High quality sequence stop: 645.
FEATURES
Location/Qualifiers
1..1070
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3890248"
/clone_lib="NIH_MGC_69"
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site_1: Not 1;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 276 a 324 c 302 g 168 t
ORIGIN
Query Match 44.0%; Score 540.8; DB 137; Length 1070;
Best Local Similarity 95.6%; Pred. No. 3.3e-127;
Matches 675; Conservative 0; Mismatches 17; Indels 14; Gaps 11;
```

```
QY 151 CGCCCCAGCGCCCTCTCCGGGACCGAAAGCGCAGCGC-CAGGGTTCTCTACCTCGA-C 208
DB 1 CGCCCCAGCGCCCTCTCCGGGACCGAAAGCGCAGCGCTCAGGGTTCTCTACCTCGATG 60
QY 209 TGCTCCGGGCCCCAG-CTGCCAGTCGAGGACCGAACCCAGCCAA-AGGGCTTCTCTTCT 266
DB 61 TGCTCCGGGCCCCAGTCTGCCAGTCGAGGAACCGAACCCAGCCAAAGAGGTTCTCTTCT 120
QY 267 GCTGC-TCACCATCTGCTTCTGCGCAGATCCTGATGGCTGAAGAGGGTGTCCCGCGCCCC 325
DB 121 GCTGCATCACCATCTGCTTCTGCGCAGATCCTGATGCTGAAGAGGGTGTCCCGCGCCCC 180
QY 326 TGCTCCAGAGGAC-GCCCTTAACGGCCGATCCCTGGC-GCCACCCCTCTGTCCCGCGT 383
DB 181 TGCTCCAGAGGACTGCCCTTAACGGCCGATCCCTGGCTGCCACCCCTCTGTGCTGCCGCT 240
QY 384 CCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGAGACCTCAGCACTTT 443
DB 241 CCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGAGACCTCAGCACTTT 300
QY 444 CCTCCAGCAACACCCCGCCCTTCTAACTGTGACTTCCCGGCACTCCCAAAAGAAATTC 503
DB 301 CCTCCAGCAACACCCCGCCCTTCTAACTGTGACTTCCCGGCACTCCCAAAAGAAATTC 360
QY 504 GAAAAACCAAGAAACACACAGCGCTACTGCTGGCGGAGAGCGTATCCCCAACCTGGGA 563
DB 361 GAAAAACCAAGAAACACACAGCGCTACTGCTGGCGGAGAGCGTATCCCCAACCTGGGA 420
QY 564 CTTCCGAGGCAACTTGAACCTCAGAACACTACACCGGAGAGCGCCACCTGGTGTGAGGCG 623
DB 421 CTTCCGAGGCAACTTGAACCTCAGAACACTACACCGGAGAGCGCCACCTGGTGTGAGGCG 480
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 22:04:48 ; Search time 2136.76 Seconds  
(without alignments)  
509,900 Million cell updates/sec

Title: US-08-799-910-9\_COPY\_211\_468  
Perfect score: 258  
Sequence: 1 GTCCGGCGCAGTCCAGT.....AGCAACACCCGGCGCCTTC 258

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 12325308 seqs, 2111498085 residues

Total number of hits satisfying chosen parameters: 24650616

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: /cgn2\_6/ptodata/2/pna/pctus\_COMB.seq.\*
  - 2: /cgn2\_6/ptodata/2/pna/US06\_COMB.seq.\*
  - 3: /cgn2\_6/ptodata/2/pna/US07\_COMB.seq.\*
  - 4: /cgn2\_6/ptodata/2/pna/US080\_COMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/pna/US081\_COMB.seq.\*
  - 6: /cgn2\_6/ptodata/2/pna/US082\_COMB.seq.\*
  - 7: /cgn2\_6/ptodata/2/pna/US083\_COMB.seq.\*
  - 8: /cgn2\_6/ptodata/2/pna/US084\_COMB.seq.\*
  - 9: /cgn2\_6/ptodata/2/pna/US085\_COMB.seq.\*
  - 10: /cgn2\_6/ptodata/2/pna/US086\_COMB.seq.\*
  - 11: /cgn2\_6/ptodata/2/pna/US087\_COMB.seq.\*
  - 12: /cgn2\_6/ptodata/2/pna/US088\_COMB.seq.\*
  - 13: /cgn2\_6/ptodata/2/pna/US089\_COMB.seq.\*
  - 14: /cgn2\_6/ptodata/2/pna/US090\_COMB.seq.\*
  - 15: /cgn2\_6/ptodata/2/pna/US091\_COMB.seq.\*
  - 16: /cgn2\_6/ptodata/2/pna/US092\_COMB.seq.\*
  - 17: /cgn2\_6/ptodata/2/pna/US093\_COMB.seq.\*
  - 18: /cgn2\_6/ptodata/2/pna/US094\_COMB.seq.\*
  - 19: /cgn2\_6/ptodata/2/pna/US095A\_COMB.seq.\*
  - 20: /cgn2\_6/ptodata/2/pna/US095B\_COMB.seq.\*
  - 21: /cgn2\_6/ptodata/2/pna/US095C\_COMB.seq.\*
  - 22: /cgn2\_6/ptodata/2/pna/US096A\_COMB.seq.\*
  - 23: /cgn2\_6/ptodata/2/pna/US096B\_COMB.seq.\*
  - 24: /cgn2\_6/ptodata/2/pna/US097\_COMB.seq.\*
  - 25: /cgn2\_6/ptodata/2/pna/US6000\_COMB.seq.\*
  - 26: /cgn2\_6/ptodata/2/pna/US6001\_COMB.seq.\*
  - 27: /cgn2\_6/ptodata/2/pna/US6002\_COMB.seq.\*
  - 28: /cgn2\_6/ptodata/2/pna/US6003\_COMB.seq.\*
  - 29: /cgn2\_6/ptodata/2/pna/US6004\_COMB.seq.\*
  - 30: /cgn2\_6/ptodata/2/pna/US6005\_COMB.seq.\*
  - 31: /cgn2\_6/ptodata/2/pna/US6006\_COMB.seq.\*
  - 32: /cgn2\_6/ptodata/2/pna/US6007\_COMB.seq.\*
  - 33: /cgn2\_6/ptodata/2/pna/US6008\_COMB.seq.\*
  - 34: /cgn2\_6/ptodata/2/pna/US6009\_COMB.seq.\*
  - 35: /cgn2\_6/ptodata/2/pna/US6010\_COMB.seq.\*
  - 36: /cgn2\_6/ptodata/2/pna/US6011\_COMB.seq.\*
  - 37: /cgn2\_6/ptodata/2/pna/US6012\_COMB.seq.\*
  - 38: /cgn2\_6/ptodata/2/pna/US6013\_COMB.seq.\*
  - 39: /cgn2\_6/ptodata/2/pna/US6014\_COMB.seq.\*
  - 40: /cgn2\_6/ptodata/2/pna/US6015\_COMB.seq.\*
  - 41: /cgn2\_6/ptodata/2/pna/US6016\_COMB.seq.\*
  - 42: /cgn2\_6/ptodata/2/pna/US6017\_COMB.seq.\*
  - 43: /cgn2\_6/ptodata/2/pna/US6018\_COMB.seq.\*

- 44: /cgn2\_6/ptodata/2/pna/US6019\_COMB.seq.\*
- 45: /cgn2\_6/ptodata/2/pna/US6020\_COMB.seq.\*
- 46: /cgn2\_6/ptodata/2/pna/US6021\_COMB.seq.\*
- 47: /cgn2\_6/ptodata/2/pna/US6022\_COMB.seq.\*
- 48: /cgn2\_6/ptodata/2/pna/US6023\_COMB.seq.\*
- 49: /cgn2\_6/ptodata/2/pna/US6024\_COMB.seq.\*
- 50: /cgn2\_6/ptodata/2/pna/US6025\_COMB.seq.\*
- 51: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB.seq.\*
- 52: /cgn2\_6/ptodata/1/pna/US05\_NEW\_COMB.seq.\*
- 53: /cgn2\_6/ptodata/1/pna/US07\_NEW\_COMB.seq.\*
- 54: /cgn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq.\*
- 55: /cgn2\_6/ptodata/1/pna/US09\_NEW\_COMB.seq.\*
- 56: /cgn2\_6/ptodata/1/pna/US60\_NEW\_COMB.seq.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	258	100.0	453	19	US-09-528-409-88514
2	258	100.0	506	24	US-09-726-805-429
3	258	100.0	521	22	US-09-644-869-356
4	258	100.0	521	44	US-60-196-718-2355
5	258	100.0	529	23	US-09-699-998-4785
6	258	100.0	617	17	US-09-396-885-430
7	258	100.0	657	19	US-09-522-303-1025
8	258	100.0	673	23	US-09-698-010-14935
9	258	100.0	673	23	US-09-699-998-9797
10	258	100.0	673	24	US-09-710-286-3772
11	258	100.0	673	24	US-09-716-473-2500
12	258	100.0	673	24	US-09-721-588-4548
13	258	100.0	673	55	US-09-726-787-2982
14	258	100.0	706	23	US-09-699-998-1778
15	258	100.0	803	23	US-09-699-998-10469
16	258	100.0	888	23	US-09-652-109-10087
17	258	100.0	888	23	US-09-652-126-9303
18	258	100.0	888	23	US-09-652-816-9065
19	258	100.0	888	24	US-09-716-473-2323
20	258	100.0	888	24	US-09-716-990-1375
21	258	100.0	888	24	US-09-721-588-4181
22	258	100.0	888	24	US-09-726-172-2642
23	258	100.0	888	24	US-09-726-805-1977
24	258	100.0	888	55	US-09-726-809-2896
25	258	100.0	1228	11	US-08-799-910-9
26	258	100.0	1228	12	US-08-825-486-9
27	258	100.0	1228	12	US-08-826-248-9
28	258	100.0	1228	12	US-08-870-434-5
29	258	100.0	1228	13	US-08-925-588-9
30	258	100.0	1228	17	US-09-372-044-9
31	258	100.0	1308	1	PCT-US00-15136-46
32	258	100.0	1316	17	US-09-338-425-2393
33	258	100.0	1316	22	US-09-649-162-9722
34	258	100.0	1316	23	US-09-652-109-9454
35	258	100.0	1316	23	US-09-652-128-10214
36	258	100.0	1316	23	US-09-652-914-9119
37	258	100.0	1316	23	US-09-699-998-9799
38	258	100.0	1316	24	US-09-710-280-2658
39	258	100.0	1316	24	US-09-716-473-2888
40	258	100.0	1316	24	US-09-721-588-3900
41	258	100.0	1316	24	US-09-726-171-2410
42	258	100.0	1316	24	US-09-726-790-2393
43	258	100.0	1316	24	US-09-726-791-1703
44	258	100.0	1316	24	US-09-726-805-1978
45	258	100.0	1316	55	US-09-726-787-2984

ALIGNMENTS



## RESULT 1

US-09-528-409-88514  
; Sequence 88514, Application US/09528409  
; GENERAL INFORMATION:  
; APPLICANT: Drmanac, Radoje T.  
; APPLICANT: Label, Ivan  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Dickson, Mark  
; APPLICANT: Jones, Lee W.  
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained  
; FILE REFERENCE: 774  
; CURRENT APPLICATION NUMBER: US/09/528,409  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/125,453  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 116231  
; SOFTWARE: Hy-patent.pl Version 3.1  
; SEQ ID NO 88514  
; LENGTH: 453  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(453)  
; OTHER INFORMATION: n = A,T,C or G

US-09-528-409-88514

Query Match 100.0%; Score 258; DB 19; Length 453;  
Best Local Similarity 100.0%; Pred. No. 7.4e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGGCGCCAGCTGCCAGTCCGAGAACCGAACCCAGCAAGGCTTCCTCTCTCTGCTG 60  
DB 51 gtcggcgccagctgccagtcgaggaacccagcaagggcttcctctctctgctg 110  
QY 61 CTCACATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 120  
DB 111 ctcaccatcgctctctccagatcctgagctgaagagggtgtgcccggccctgcct 170  
QY 121 CCAGAGAGCCCTTACGCGCGATCCTGCGGCCACCCCTGTGTCCCGCTCTCGAG 180  
DB 171 ccagagagcccttaacgcgcgcatccctggtgcccaccctgtgccccgcctcgag 230  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACCTTCCTCCAG 240  
DB 231 ccccttaatctgacttcggagccctcggaactacgctctggacctcagcacttccctcag 290  
QY 241 CAACACCCGCGCCCTTC 258  
DB 291 caacacccgcgcccttc 308

## RESULT 2

US-09-726-805-429  
; Sequence 429, Application US/09726805  
; GENERAL INFORMATION:  
; APPLICANT: Gearing, David P.  
; APPLICANT: Hollzman, Douglas A.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; FILE REFERENCE: 1600.2017-001  
; CURRENT APPLICATION NUMBER: US/09/726,805  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: 60/168,140  
; PRIOR FILING DATE: 1999-11-30  
; NUMBER OF SEQ ID NOS: 2158  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 429  
; LENGTH: 506  
; TYPE: DNA  
; ORGANISM: Homo sapiens

US-09-726-805-429

Query Match 100.0%; Score 258; DB 24; Length 506;  
Best Local Similarity 100.0%; Pred. No. 7.5e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGGCGCCAGCTGCCAGTCCGAGAACCGAACCCAGCAAGGCTTCCTCTCTCTGCTG 60  
DB 138 gtcggcgccagctgccagtcgaggaacccagcaagggcttcctctctctgctg 197  
QY 61 CTCACATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 120  
DB 198 ctcaccatcgctctctccagatcctgagctgaagagggtgtgcccggccctgcct 257  
QY 121 CCAGAGAGCCCTTACGCGCGATCCTGCGGCCACCCCTGTGTCCCGCTCTCGAG 180  
DB 258 ccagagagcccttaacgcgcgcatccctggtgcccaccctgtgccccgcctcgag 317  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACCTTCCTCCAG 240  
DB 318 ccccttaatctgacttcggagccctcggaactacgctctggacctcagcacttccctcag 377  
QY 241 CAACACCCGCGCCCTTC 258  
DB 378 caacacccgcgcccttc 395

## RESULT 3

US-09-644-869-356  
; Sequence 356, Application US/09644869  
; GENERAL INFORMATION:  
; APPLICANT: Shyjan, Andrew W.  
; APPLICANT: McCarthy, Sean A.  
; APPLICANT: Holtzman, Douglas A.  
; APPLICANT: Monehan, John  
; APPLICANT: Richardson, Jennifer  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; FILE REFERENCE: 1600.1182-001  
; CURRENT APPLICATION NUMBER: US/09/644,869  
; PRIOR FILING DATE: 2000-08-28  
; PRIOR APPLICATION NUMBER: 60/151,062  
; PRIOR FILING DATE: 1999-08-27  
; NUMBER OF SEQ ID NOS: 9708  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 356  
; LENGTH: 521  
; TYPE: DNA  
; ORGANISM: Homo sapiens

US-09-644-869-356

Query Match 100.0%; Score 258; DB 22; Length 521;  
Best Local Similarity 100.0%; Pred. No. 7.5e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGCGCCAGCTGCCAGTCCGAGAACCGAACCCAGCAAGGCTTCCTCTCTCTGCTG 60  
DB 211 gtcggcgccagctgccagtcgaggaacccagcaagggcttcctctctctgctg 270  
QY 61 CTCACATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 120  
DB 271 ctcaccatcgctctctccagatcctgagctgaagagggtgtgcccggccctgcct 330  
QY 121 CCAGAGAGCCCTTACGCGCGATCCTGCGGCCACCCCTGTGTCCCGCTCTCGAG 180  
DB 331 ccagagagcccttaacgcgcgcatccctggtgcccaccctgtgccccgcctcgag 390  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACCTTCCTCCAG 240  
DB 391 ccccttaatctgacttcggagccctcggaactacgctctggacctcagcacttccctcag 450



QY 241 CAACACCGGCCCGCCTTC 258  
|||||  
Db 346 caacacccggccgccttc 363

## RESULT 7

```

US-09-522-303-1025
; Sequence 1025, Application US/09522303
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: HUMAN MAMMARY EPITHELIAL CELL
; FILE REFERENCE: 1600.1086-001
; CURRENT APPLICATION NUMBER: US/09/522
; CURRENT FILING DATE: 2000-03-08
; EARLIER APPLICATION NUMBER: 60/123,399
; EARLIER FILING DATE: 1999-03-08
; NUMBER OF SEQ ID NOS: 1353
; SOFTWARE: FastSeq for Windows Version
; SEQ ID NO 1025
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-522-303-1026

```

Query Match	100.0%	Score 258;	DB 19;	Length 657;
Best Local Similarity	100.0%			
Matches 258; Conservative	0;	Mismatches	0;	
		Pred. No.	7.6e-52;	

Qy	1	GTCCGGCCAGTGCAGTCGAGGAAACGAACCCAGCCAAAGGCTTCTCTTCTCTGCTG	60
Db	277	gtcggcgccagctgcagtcgaggaacgaacccaaaggctctcttctctgctg	336
Qy	61	CTACCAATCTCTCTGCCAGATCTTGCTGAAGAGGGTGTCGCGCGGCCCTCGCT	120
Db	337	ctacccatcgctctcgcagatacctgtaggtgaagggtgtagcgcgcctcgctc	396
Qy	121	CCAGAGACGCCCTTAACGCCGATCCTCTGGCGCCACCCCTGTGTCCCGCTCTCGAG	180
Db	397	ccagagagcgccctaaegccgcatccctcggcgccacccctcgtgcccccgcctcgag	456
Qy	181	CCCTTTTAATCTGACTTCTGGAGCGCCCGGACTACGCTCTGGACCTCAGCACTTCTCCAG	240
Db	457	ccctttaaactgacttcggagccctcggactacgtctggacctcagcacttctctccag	516
Qy	241	CAACACCCGGCGCCCTTC	258
Db	517	caacacccggccgccttc	534

## RESULT 8

```

US-09-698-010-14935
:
: Sequence 14935, Application US/09698010
: GENERAL INFORMATION:
: APPLICANT: Williamson, Mark
: APPLICANT: Shyjan, Andrew W.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: TITLE OF INVENTION: THEREFOR
: FILE REFERENCE: 1600.2029-001
: CURRENT APPLICATION NUMBER: US/09/698,010
: CURRENT FILING DATE: 2000-10-27
: PRIOR APPLICATION NUMBER: 60/162,358
: PRIOR FILING DATE: 1999-10-29
: NUMBER OF SEQ ID NOS: 15684
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 14935
: LENGTH: 673
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-698-010-14935

```

US-09-698-010-14935

Query Match	100.0%	Score 258;	DB 23;	Length 673;
Best Local Similarity	100.0%;	Pred. No. 7.7e-12;		
Matches 258; Conservative	0;	Mismatches	0;	Indels
	0;	Gaps		
QY 1	GTCCGGCGCCAGCTGCCAGTTCGAGGAACCGAACCCAGCCAAAGAGCTTCTCTTTCTGCTG	60		
Db 269	gtccgcgccagctgccagtcgaggaacgcgacccagccaaagactctctcttctgclg	328		
QY 61	CTCACCATCGTCTTCTGCGAGATCTGTATGGCTGAAGAGGCTGTGCGCGCGCCCTTGCCT	120		
Db 329	ctcaccatcgtcttctgcagatcctgctgagagggtgtgcgcgcgcctgcct	398		
QY 121	CCAGGAGCGCCCTTAAGCGCGCATCTCTGCGCGCCACCCCTGTGTCCTCCGCTCTCGAG	180		
Db 389	ccagaggacgcgccctaaacgcgcgcatcctcgcgccacccctctgtctccccgtccctcgaq	448		
QY 181	CCCTTTAATCTGCATTCGAGGCCCTCGGACTTACGCTCTGGACCTCAGCACTTCTCCACG	240		
Db 449	ccctttaactcgactctcgagccctcagactacgctctggacctcagcactttctccacg	508		
QY 241	CAACACCGGCGCGCTTC	258		
Db 509	caacacccggcgccctc	526		

## RESULT 9

```

US-09-699-998-9797
; Sequence 9797, Application US/09699998
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gearing, David P.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2008-001
; CURRENT APPLICATION NUMBER: US/09/699,998
; CURRENT FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/162,362
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 10905
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9797
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-699-998-9797

```

Query Match	Score 258;	DB 23;	Length 673;
Best Local similarity	100.0%		
Best Local similarity	100.0%		

Matches	258	Conservative	0	Mismatches	0	Indels	0	Gaps
Qy	1	FTCCGGCGCCAGCTGCCAGTCGAGCAACCGAACCCAGCCAAAGAGCTTCTCTTCTGCG	60					
Db	269	gtccggcgccagctcgaagtcgaaggaaacccagcaaaagctctctttctgtg	328					
Qy	61	CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGCCT	120					
Db	329	ctcaccatcgtctctctgcagatcctgatgctgaagagggtgtgcgcggcgccctgcct	388					
Qy	121	CCAGGAGGACGCCCTAACGCCGATTCCTGTGGCGCCCACCCCTGTGTCCCGCGTCTCCAG	180					
Db	389	ccagaggacgcccctaatacgcgcatacctcgtgcgcacccctcgtgtccccgcctccag	448					
Qy	181	CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTTCAGACCTTCCTCCAG	240					
Db	449	ccctttaatctgacttcgagccctcggactaagctctgacctcagcactttctccag	508					
Qy	241	CAACACCCGGCGGCTTTC	258					
Db	509	caacacccggcgcccttc	526					

Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10  
US-09-710-286-3772  
; Sequence 3772, Application US/09710286  
; GENERAL INFORMATION:  
; APPLICANT: Gearling, David P.  
; APPLICANT: McCarthy, Sean A.  
; APPLICANT: Holtzman, Douglas A.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; TITLE OF INVENTION: THEREFOR  
; FILE REFERENCE: 1600.2005-001  
; CURRENT APPLICATION NUMBER: US/09/710,286  
; CURRENT FILING DATE: 2000-11-10  
; PRIOR APPLICATION NUMBER: 60/164,255  
; PRIOR FILING DATE: 1999-11-04  
; NUMBER OF SEQ ID NOS: 4115  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3772  
; LENGTH: 673  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-710-286-3772

Query Match 100.0%; Score 258; DB 24; Length 673;  
Best Local Similarity 100.0%; Pred. No. 7.7e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGGCGCCAGCTGCCAGTCCGAGAACCCAGCCAAAGGCTTCCTCTCTGCTG 60  
DB 269 gtccggcgccagctgccagtcgaggaacccgacccagccaaagagctctctcttctgctg 328  
QY 61 CTCACCATCGTCTTCTGCCAGATCCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCT 120  
DB 329 ctacccatcgctctctgccagatccctgagagaggggtgctgcggcgccctgctcct 388  
QY 121 CCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGCTGTCCTCCCGGCTCGAG 180  
DB 389 ccagagagacgcccctaacgcccgcctccctggcgccccccctgctgccccgctcctcag 448  
QY 181 CCCTTTAATCTGACTTGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCTCCAG 240  
DB 449 ccccttaactgacttcggagccctcgactacgctctggacctcagcacttctctccag 508  
QY 241 CAACACCGCGCCGCTTC 258  
DB 509 caaccccgccgccccttc 526

RESULT 12  
US-09-721-588-4548  
; Sequence 4548, Application US/09721588  
; GENERAL INFORMATION:  
; APPLICANT: Gearling, David P.  
; APPLICANT: Holtzman, Douglas A.  
; APPLICANT: Villevall, Jean-Luc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; TITLE OF INVENTION: THEREFOR  
; FILE REFERENCE: 1600.2046-001  
; CURRENT APPLICATION NUMBER: US/09/721,588  
; CURRENT FILING DATE: 2000-11-22  
; PRIOR APPLICATION NUMBER: 60/167,381  
; PRIOR FILING DATE: 1999-11-24  
; NUMBER OF SEQ ID NOS: 5410  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4548  
; LENGTH: 673  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-721-588-4548

Query Match 100.0%; Score 258; DB 24; Length 673;  
Best Local Similarity 100.0%; Pred. No. 7.7e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGGCGCCAGCTGCCAGTCCGAGAACCCAGCCAAAGGCTTCCTCTCTGCTG 60  
DB 269 gtccggcgccagctgccagtcgaggaacccgacccagccaaagagctctctcttctgctg 328  
QY 61 CTCACCATCGTCTTCTGCCAGATCCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCT 120  
DB 329 ctacccatcgctctctgccagatccctgagagaggggtgctgcggcgccctgctcct 388  
QY 121 CCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGCTGTCCTCCCGGCTCGAG 180  
DB 389 ccagagagacgcccctaacgcccgcctccctggcgccccccctgctgccccgctcctcag 448  
QY 181 CCCTTTAATCTGACTTGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCTCCAG 240  
DB 449 ccccttaactgacttcggagccctcgactacgctctggacctcagcacttctctccag 508  
QY 241 CAACACCGCGCCGCTTC 258  
DB 509 caaccccgccgccccttc 526

RESULT 13  
US-09-726-787-2982

Query Match 100.0%; Score 258; DB 24; Length 673;  
Best Local Similarity 100.0%; Pred. No. 7.7e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGGCGCCAGCTGCCAGTCCGAGAACCCAGCCAAAGGCTTCCTCTCTGCTG 60  
DB 269 gtccggcgccagctgccagtcgaggaacccgacccagccaaagagctctcttctgctg 328  
QY 61 CTCACCATCGTCTTCTGCCAGATCCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCT 120  
DB 329 ctacccatcgctctctgccagatccctgagagaggggtgctgcggcgccctgctcct 388  
QY 121 CCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGCTGTCCTCCCGGCTCGAG 180  
DB 389 ccagagagacgcccctaacgcccgcctccctggcgccccccctgctgccccgctcctcag 448  
QY 181 CCCTTTAATCTGACTTGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCTCCAG 240  
DB 449 ccccttaactgacttcggagccctcgactacgctctggacctcagcacttctctccag 508  
QY 241 CAACACCGCGCCGCTTC 258  
DB 509 caaccccgccgccccttc 526

Query Match 100.0%; Score 258; DB 24; Length 673;  
Best Local Similarity 100.0%; Pred. No. 7.7e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGGCGCCAGCTGCCAGTCCGAGAACCCAGCCAAAGGCTTCCTCTCTGCTG 60  
DB 269 gtccggcgccagctgccagtcgaggaacccgacccagccaaagagctctcttctgctg 328  
QY 61 CTCACCATCGTCTTCTGCCAGATCCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCT 120  
DB 329 ctacccatcgctctctgccagatccctgagagaggggtgctgcggcgccctgctcct 388  
QY 121 CCAGAGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGCTGTCCTCCCGGCTCGAG 180  
DB 389 ccagagagacgcccctaacgcccgcctccctggcgccccccctgctgccccgctcctcag 448  
QY 181 CCCTTTAATCTGACTTGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCTCCAG 240  
DB 449 ccccttaactgacttcggagccctcgactacgctctggacctcagcacttctctccag 508  
QY 241 CAACACCGCGCCGCTTC 258  
DB 509 caaccccgccgccccttc 526

RESULT 11  
US-09-716-473-2500  
; Sequence 2500, Application US/09716473  
; GENERAL INFORMATION:  
; APPLICANT: Williamson, Mark  
; APPLICANT: Shyjan, Andrew W.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; TITLE OF INVENTION: THEREFOR  
; FILE REFERENCE: 1600.2043-001  
; CURRENT APPLICATION NUMBER: US/09/716,473  
; CURRENT FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: 60/166,502  
; PRIOR FILING DATE: 1999-11-19  
; NUMBER OF SEQ ID NOS: 2933  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2500  
; LENGTH: 673  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-716-473-2500

Query Match 100.0%; Score 258; DB 24; Length 673;  
Best Local Similarity 100.0%; Pred. No. 7.7e-52;

; Sequence 2982, Application US/09726787  
; GENERAL INFORMATION:  
; APPLICANT: Gearing, David P.  
; APPLICANT: Holtzman, Douglas A.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; TITLE OF INVENTION: THEREFOR  
; FILE REFERENCE: 1600.2010-001  
; CURRENT APPLICATION NUMBER: US/09/726,787  
; CURRENT FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: 60/168,132  
; PRIOR FILING DATE: 1999-11-30  
; NUMBER OF SEQ ID NOS: 3241  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2982  
; LENGTH: 673  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-726-787-2982

Query Match 100.0%; Score 258; DB 55; Length 673;  
Best Local Similarity 100.0%; Pred. No. 7.7e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCCGGCCAGCTGCCAGTCCAGAACCCGACCCGACCAAGGCTTCTCTTCTGCTG 60  
Db 269 gtccggccagctgccagtcagagaaacccagccaaagagcttcttctgctg 328  
Qy 61 CTCACCATCTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCGCTGCT 120  
Db 329 ctcaccatctcttctgccagatcctgatgctgaagaggtgtgccggccgctgct 388  
Qy 121 CCAGAGGACGCCCTAACCGCATCCCTGGCGCCACCCCTGTGTCCTCCGCTCGAG 180  
Db 389 ccagaggacgcccctaaacgcgcctccctggcgccacccctgtgcccgctcctgag 448  
Qy 181 CCCTTTAATCTGACTTCGGACCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCCAG 240  
Db 449 ccctttaatctgacttcggacccctcggaactacgctctggacctcagcacttctccag 508  
Qy 241 CAACACCGCGCGCTTC 258  
Db 509 caacaccggcgcccttc 526

RESULT 14  
US-09-699-998-1778  
; Sequence 1778, Application US/09699998  
; GENERAL INFORMATION:  
; APPLICANT: Holtzman, Douglas A.  
; APPLICANT: Gearing, David P.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; TITLE OF INVENTION: THEREFOR  
; FILE REFERENCE: 1600.2008-001  
; CURRENT APPLICATION NUMBER: US/09/699,998  
; CURRENT FILING DATE: 2000-10-30  
; PRIOR APPLICATION NUMBER: 60/162,362  
; PRIOR FILING DATE: 1999-10-29  
; NUMBER OF SEQ ID NOS: 10905  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1778  
; LENGTH: 706  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(706)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-699-998-1778

Query Match 100.0%; Score 258; DB 23; Length 706;  
Best Local Similarity 100.0%; Pred. No. 7.7e-52;

Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GTCCGGCCAGCTGCCAGTCCAGAACCCGACCCGACCAAGGCTTCTCTTCTGCTG 60  
Db 195 gtccggccagctgccagtcagagaaacccagccaaagagcttcttctgctg 254  
Qy 61 CTCACCATCTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCGCTGCT 120  
Db 255 ctcaccatctcttctgccagatcctgatgctgaagaggtgtgccggccgctgct 314  
Qy 121 CCAGAGGACGCCCTAACCGCATCCCTGGCGCCACCCCTGTGTCCTCCGCTCGAG 180  
Db 315 ccagaggacgcccctaaacgcgcctccctggcgccacccctgtgcccgctcctgag 374  
Qy 181 CCCTTTAATCTGACTTCGGACCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCCAG 240  
Db 375 ccctttaatctgacttcggacccctcggaactacgctctggacctcagcacttctccag 434  
Qy 241 CAACACCGCGCGCTTC 258  
Db 435 caacaccggcgcccttc 452

RESULT 15  
US-09-699-998-10469  
; Sequence 10469, Application US/09699998  
; GENERAL INFORMATION:  
; APPLICANT: Holtzman, Douglas A.  
; APPLICANT: Gearing, David P.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; TITLE OF INVENTION: THEREFOR  
; FILE REFERENCE: 1600.2008-001  
; CURRENT APPLICATION NUMBER: US/09/699,998  
; CURRENT FILING DATE: 2000-10-30  
; PRIOR APPLICATION NUMBER: 60/162,362  
; PRIOR FILING DATE: 1999-10-29  
; NUMBER OF SEQ ID NOS: 10905  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10469  
; LENGTH: 803  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-699-998-10469

Query Match 100.0%; Score 258; DB 23; Length 803;  
Best Local Similarity 100.0%; Pred. No. 7.8e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCCGGCCAGCTGCCAGTCCAGAACCCGACCCGACCAAGGCTTCTCTTCTGCTG 60  
Db 184 gtccggccagctgccagtcagagaaacccagccaaagagcttcttctgctg 243  
Qy 61 CTCACCATCTCTTCTGCCAGATCCTGATGCTGAAGAGGTGTGCCGGCCGCTGCT 120  
Db 244 ctcaccatctcttctgccagatcctgatgctgaagaggtgtgccggccgctcct 303  
Qy 121 CCAGAGGACGCCCTAACCGCATCCCTGGCGCCACCCCTGTGTCCTCCGCTCGAG 180  
Db 304 ccagaggacgcccctaaacgcgcctccctggcgccacccctgtgtccccgctcctgag 363  
Qy 181 CCCTTTAATCTGACTTCGGACCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCCAG 240  
Db 364 ccctttaatctgacttcggacccctcggaactacgctctggacctcagcacttctccag 423  
Qy 241 CAACACCGCGCGCTTC 258  
Db 424 caacaccggcgcccttc 441

Search completed: January 30, 2001, 22:04:50  
Job time: 13154 sec

ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5876949ris  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455.073A  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: DeLuca, Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: UPN-2201  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4362 base pairs  
TYPE: nucleic acid  
STRADEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 220..2118  
US-08-455-073A-1

Query Match 12.4%; Score 32; DB 2; Length 4362;  
Best Local Similarity 49.5%; Pred. No. 4.2;  
Matches 109; Conservative 0; Mismatches 110; Indels 1; Gaps 1;  
QY 28 CGGAACCCAGCCAAAAGGCTTCTTTCTGCTGCTCACCATCTCTTCTGCCAGATCCG 87  
DB 246 CCGCACTTCCACCACCAGCTCCCATCTTCTTTCAGCCCTAGCCGCGGAGCCCG 187  
QY 88 ATGGCTCAAGAGGCTGTGCGGCGCCCTGCTCCAGAGAGAGCCCTTAACGCGGATCC 147  
DB 186 CCCCCAGAGGTGGGTGGGCGCTCGAGGC-CCAGCGCGCGCGCGCGCGCGCGCC 128  
QY 148 CTGCGCCGCCACCCCTGTGTCGCCCGCTCGAGCCCTTTAATCTGACTTCGAGCCCTCG 207  
DB 127 GCGGCTTCCG 68  
QY 208 GACTAGCTGTGACCTCAGCACTTTTCCTCCAGCAACACC 247  
DB 67 GCGGCGCGCTGCGGAGCGCCCTTGGCAGCGCGCGCTCC 28

RESULT 15  
US-08-584-226-3/c  
Sequence 3, Application US/08584226  
Patent No. 5798240  
GENERAL INFORMATION:  
APPLICANT: Martinis, Susan A.  
APPLICANT: Sassanfar, Mandana  
APPLICANT: Kim, Sunghoon  
APPLICANT: Lee, Sang Ho  
APPLICANT: Schimmel, Paul R.  
TITLE OF INVENTION: RECOMBINANT MYCOBACTERIAL METHIONYL-tRNA  
NUMBER OF INVENTION: SYNTHETASE GENES, TESTER STRAINS AND ASSAYS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington

STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173-4799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/584,226  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/305,766  
FILING DATE: 13-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: CPI94-052  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 599 base pairs  
TYPE: nucleic acid  
STRADEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-584-226-3

Query Match 12.3%; Score 31.8; DB 1; Length 599;  
Best Local Similarity 48.1%; Pred. No. 3;  
Matches 90; Conservative 0; Mismatches 97; Indels 0; Gaps 0;  
QY 65 CCATGCTCTTCTGCCAGATCCGTGCTGAAGAGGGTGTGCGCGCGCCCTGCTGTCCTCTCGAGCCCT 184  
DB 591 CCGCGAATCTATCTGACGGGACGGATGAGCACGCCCTGAAGTGGCGGACGCGCG 532  
QY 125 AGGAGCGCCCTTAACGCGCGCATCCCTGGCGCCACCCCTGTGTCCTCTCTCGAGCCCT 184  
DB 531 CCGCGCGCGCGCTGCGGACGAGCTTCCCGGGGCAATTCGACGCTTTCACCGCA 472  
QY 185 TTAATCTGACTTCGGAGCCCTCGGACTAGCTTGTGACCTCAGCACTTTCTCTCCAGCAAC 244  
DB 471 TCGAGGAGCGCTGAACATCTCTCTCGACCGGTTTCATCGGCACACCGACCCCGACCT 412  
QY 245 ACCCGGC 251  
DB 411 ACGCGGC 405

Search completed: January 30, 2001, 21:28:34  
Job time: 19996 sec



## RESULT 10

US-08-859-201-1/c  
; Sequence 1, Application US/08859201  
; Patent No. 5880261  
; GENERAL INFORMATION:  
; APPLICANT: Waerber, G  
; APPLICANT: Nicod, P  
; TITLE OF INVENTION: Transcription Factor Islet-Brain 1 (IB1)  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Luann Cserf  
; STREET: 750 Arimo Avenue  
; CITY: Oakland  
; STATE: California  
; COUNTRY: USA  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/859,201  
; FILING DATE: 20-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9706731.8  
; FILING DATE: 03-APR-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9709920.4  
; FILING DATE: 15-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Luann Cserf  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: ME A9702  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2953 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 108..2252  
US-08-859-201-1

Query Match 13.5%; Score 34.8; DB 2; Length 2953;  
Best Local Similarity 62.8%; Pred. No. 0.69; Mismatches 32; Indels 0; Gaps 0;  
Matches 54; Conservative 0;  
QY 98 AGCGTGTCCGGCGCCCTCCAGAGAGCCCTTAACGCCGATCCCTGGGCCCA 157  
DB 2539 AGATGTGTGAGCGGCTCTGCTAGGGAGCCCTTTGTGGCCAGATCCCTGACCCCA 2480  
QY 158 CCCTGTGTCCCGCCCTCCAGAGCC 183  
DB 2479 CCCCAGCCCTGCCCTTCCAGCCCC 2454

## RESULT 11

US-08-188-582-8/c  
; Sequence 8, Application US/08188582  
; Patent No. 5534410  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlacht, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Naoko  
; APPLICANT: Wang, Edith

; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/188,582  
; FILING DATE: 28-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1120 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 80...913  
US-08-188-582-8

Query Match 13.0%; Score 33.6; DB 1; Length 1120;  
Best Local Similarity 51.0%; Pred. No. 1.2;  
Matches 104; Conservative 0; Mismatches 99; Indels 1; Gaps 1;  
QY 34 CCAGCCAAAGGCTTCTTTCTGCTGCATCCATCGTCTTCTGCAGATCCTGATGCT 93  
DB 866 CCACCTCCAAAGGCTTCTTTCTGCTGCATCCATCGTCTTCTGCAGATCCTGATGCT 93  
QY 94 GAAGAGGTGTCCCGCGCCCTGCTCCAGAGAGCGCCCTTAACGCCGATCCCTGGCG 153  
DB 807 GAACCGATGCCCGAGCGATGCTGCCACGCCCGCGATCGCTGCCACCTCCATCTTG 748  
QY 154 CCCACCTGTGTCCCGCGCTCTCGAGCGCTTAACTGCTGACITCGAGGCCCTCGGACTAC 213  
DB 747 AGATCTCGCGCGCGCTGCTCTGCTCTTAACCTCTGACACCCGCCCGCCAGAT 688  
QY 214 GCTCTGGACCTCAGCACTTTCTCTC 237  
DB 687 CCGCGGAGTGCTGCCACCGCTTTTC 664

## RESULT 12

US-08-646-715-8/c  
; Sequence 8, Application US/08646715  
; Patent No. 5637686  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlacht, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Naoko  
; APPLICANT: Wang, Edith



[illegible]



QY 181 CCCTTAAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCCTTCAGCAGCTTTCCTCCAG 240  
Db 391 ccccttaactgacttcgagccctcgactacgctctgagcccttcctccag 450  
QY 241 CAACACCGCGCGCCTTC 258  
Db 451 caacacccgcccgccttc 468

RESULT 4  
US-08-232-463-14  
: Sequence 14, Application US/08232463  
: Patent No. 5670367  
: GENERAL INFORMATION:  
: APPLICANT: DORNER, F.  
: APPLICANT: SCHEIFLINGER, F.  
: APPLICANT: FALKNER, F. G.  
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
: NUMBER OF SEQUENCES: 52  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Foley & Lardner  
: STREET: 1800 Diagonal Road, Suite 500  
: CITY: Alexandria  
: STATE: VA  
: COUNTRY: USA  
: ZIP: 22313-0299  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: APPLICATION DATA:  
: FILING DATE: US/08/232,463  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US/07/935,313  
: FILING DATE:  
: APPLICATION NUMBER: EP 91 114 300.6  
: FILING DATE: 26-AUG-1991  
: ATTORNEY/AGENT INFORMATION:  
: NAME: BENT, Stephen A.  
: REGISTRATION NUMBER: 29,768  
: REFERENCE/DOCKET NUMBER: 30472/114 1MMU  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (703)836-9300  
: TELEFAX: (703)683-4109  
: TELEX: 899149  
: INFORMATION FOR SEQ ID NO: 14:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 7218 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: IMMEDIATE SOURCE:  
: CLONE: pTZgpt-F1s  
US-08-232-463-14

Query Match 16.7%; Score 43; DB 1; Length 7218;  
Best Local Similarity 7.4%; Pred. No. 0.0054;  
Matches 19; Conservative 139; Mismatches 99; Indels 0; Gaps 0;  
QY 2 TCCGGCCGCGCCTCCAGTCGAGGAGCGAACCAGCGCAAAAGCTTCTCTTCCTGCTGC 61  
Db 1032 TCCGAGCTGCTCGAGTCGAGTCGAGGAGTTCGATGCTGAAGAGGTCGCGCGCCCTGCTC 121  
QY 62 TCACATCGCTCTTCGAGAGTTCGATGCTGAAGAGGTCGCGCGCCCTGCTC 121  
Db 1052 YY 1151  
QY 122 CAGAGGACGCCCTTAACGCCCATCCCTGCGCGCCACCCCTGTCTCCCTCCCTCCAGC 181

Db 1152 YY 1211  
QY 182 CCTTAAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCCTTCAGCAGCTTTCCTCCAGC 241  
Db 1212 YY 1271  
QY 242 AACACCGCGCGCCTTC 258  
Db 1272 YYYYYYYYYYYYYYYYYY 1288

RESULT 5  
US-08-087-007-2  
: Sequence 2, Application US/08087007  
: Patent No. 5705732  
: Patent No. 5705732 5684223  
: GENERAL INFORMATION:  
: APPLICANT: Sims, Peter J.  
: APPLICANT: Bothwell, Alfred L.M.  
: APPLICANT: Elliott, Eileen A.  
: APPLICANT: Flavell, Richard A.  
: APPLICANT: Madri, Joseph  
: APPLICANT: Rollins, Scott  
: APPLICANT: Bell, Leonard  
: APPLICANT: Squinto, Stephen  
: TITLE OF INVENTION: Universal Donor Cells  
: NUMBER OF SEQUENCES: 4  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Kilpatrick & Cody  
: STREET: 1100 Peachtree Street, Suite 2800  
: CITY: Atlanta  
: STATE: Georgia  
: COUNTRY: U.S.  
: ZIP: 30309-4530  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: APPLICATION DATA:  
: FILING DATE: US/08/087,007  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Fabst, Patrea L.  
: REGISTRATION NUMBER: 31,284  
: REFERENCE/DOCKET NUMBER: OMRF135  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 404-815-6500  
: TELEFAX: 404-815-6555  
: INFORMATION FOR SEQ ID NO: 2:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 2847 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: cDNA  
: HYPOTHETICAL: NO  
: ANTI-SENSE: NO  
: ORIGINAL SOURCE:  
: ORGANISM: Homo sapiens  
: IMMEDIATE SOURCE:  
: LIBRARY: GenBank HUMDAF; HUMDAF1  
: CLONE: Human DAF cDNA  
: FEATURE:  
: NAME/KEY: misc-feature  
: LOCATION: 1..819  
: OTHER INFORMATION: /note= "HUMDAF1 (Promotor and 5'  
: OTHER INFORMATION: end of Exon 1, genomic sequence)"  
US-08-087-007-2

Query Match 15.3%; Score 39.4; DB 1; Length 2847;

US-08-826-246-9

Query Match 100.0%; Score 258; DB 3; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 1.2e-60;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCCGGCGCAGCTGCCATTCGAGGAAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
Db 211 GTCCGGCGCAGCTGCCATTCGAGGAAACCCAGCCAAAGGCTTCTTTCTGCTG 270  
OY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCCCTGCT 120  
Db 271 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCCCTGCT 330  
OY 121 CCAGAGGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGCTCCCGCTCCTGAG 180  
Db 331 CCAGAGGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGCTCCCGCTCCTGAG 390  
OY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACCTTCTCTCCAG 240  
Db 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACCTTCTCTCCAG 450  
OY 241 CAACACCGCGCCCTTC 258  
Db 451 CAACACCGCGCCCTTC 468

## RESULT 2

US-08-944-495-9  
; Sequence 9, Application US/08944495  
; Patent No. 6087477

## GENERAL INFORMATION:

APPLICANT: Falb, Dean  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
TREATMENT AND DIAGNOSIS OF  
TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESSES:

ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/944,495

## CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/799,910  
FILING DATE:

## ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-067-999  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
TELEX: 66141 PENNIE

## INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1228 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:

NAME/KEY: Coding Sequence  
LOCATION: 1...468  
OTHER INFORMATION:  
US-08-944-495-9

Query Match 100.0%; Score 258; DB 3; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 1.2e-60;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCCGGCGCAGCTGCCAGTCCAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
Db 211 GTCCGGCGCAGCTGCCAGTCCAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 270  
OY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCCCTGCT 120  
Db 271 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCCCTGCT 330  
OY 121 CCAGAGGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGCTCCCGCTCCTGAG 180  
Db 331 CCAGAGGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGCTCCCGCTCCTGAG 390  
OY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACCTTCTCTCCAG 240  
Db 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACCTTCTCTCCAG 450  
OY 241 CAACACCGCGCCCTTC 258  
Db 451 CAACACCGCGCCCTTC 468

## RESULT 3

US-09-126-640-5  
; Sequence 5, Application US/09126640A  
; Patent No. 6099823

## GENERAL INFORMATION:

APPLICANT: Falb, Dean A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE  
FILE REFERENCE: 7853-126  
CURRENT APPLICATION NUMBER: US/09/126,640A

CURRENT FILING DATE: 1998-07-30

EARLIER APPLICATION NUMBER: 08/870,414

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 08/799,910

EARLIER FILING DATE: 1997-02-13

EARLIER APPLICATION NUMBER: 60/011,787

EARLIER FILING DATE: 1996-02-16

NUMBER OF SEQ ID NOS: 44

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5

LENGTH: 1228

TYPE: DNA

ORGANISM: Homo sapiens

US-09-126-640-5

Query Match 100.0%; Score 258; DB 3; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 1.2e-60;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCCGGCGCAGCTGCCAGTCCAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
Db 211 GTCCGGCGCAGCTGCCAGTCCAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 270  
OY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCCCTGCT 120  
Db 271 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCCCTGCT 330  
OY 121 CCAGAGGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGCTCCCGCTCCTGAG 180  
Db 331 CCAGAGGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGCTCCCGCTCCTGAG 390

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 21:28:30 ; Search time 49.1 Seconds  
(without alignments)  
846.830 Million cell updates/sec

Title: US-08-799-910-9\_COPY\_211\_468  
Perfect score: 258  
Sequence: 1 GTCGGCGCCACTGCCAGT.....ACCAACACCCGCCGCCCTTC 258

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	258	100.0	1228	3	US-08-826-246-9
2	258	100.0	1228	3	US-08-944-495-9
3	258	100.0	1228	3	US-09-126-640-5
4	39.4	15.3	2847	1	US-08-232-463-14
5	39.4	15.3	2847	1	US-08-087-007-2
6	39.4	15.3	2847	3	US-08-483-433-2
7	39.4	15.3	2847	4	PCT-US92-05920-2
8	38.6	15.0	6354	3	US-09-058-389A-5
9	37.6	14.6	170	3	US-09-058-389A-13
10	34.8	13.5	2953	2	US-08-859-201-1
11	33.6	13.0	1120	1	US-08-188-582-8
12	33.6	13.0	1120	1	US-08-646-715-8
13	32	12.4	3765	3	US-07-705-490-1
14	32	12.4	4362	2	US-08-455-073A-1
15	31.8	12.3	599	1	US-08-584-226-3
16	31.6	12.2	2074	2	US-09-018-576-2
17	31.6	12.2	2074	2	US-09-018-576-11
18	31.6	12.2	2074	3	US-09-248-137-2
19	31.6	12.2	2074	3	US-09-248-137-11
20	31.4	12.2	1335	4	PCT-US91-06532-1
21	31.4	12.2	38506	3	US-09-320-878-19
22	30.8	11.9	797	1	US-08-332-467B-2
23	30.8	11.9	797	1	US-08-681-811-2
24	30.8	11.9	797	4	PCT-US93-12507-2
25	30.8	11.9	2680	3	US-09-063-035-1
26	30.6	11.9	7791	3	US-08-949-386-23
27	30.6	11.9	7791	3	US-08-450-562-23
28	30.6	11.9	7808	3	US-08-949-386-22

c	29	30.6	11.9	7808	3	US-08-450-562-22
	30	30.2	11.7	22481	4	PCT-US95-07201-43
	31	30	11.6	3688	5	5248670-4
	32	29.8	11.6	536	1	US-08-329-704-1
	33	29.8	11.6	536	2	US-08-472-604-1
	34	29.8	11.6	536	2	US-08-486-117-1
	35	29.8	11.6	536	2	US-08-865-337A-3
	36	29.8	11.6	2196	2	US-08-865-337A-2
	37	29.8	11.6	8299	1	US-08-462-014-2
	38	29.8	11.6	8299	3	US-08-923-137-3
	39	29.6	11.5	693	1	US-08-168-091A-3
c	40	29.6	11.5	1219	4	PCT-US93-06251-11
	41	29.6	11.5	1313	1	US-08-176-427B-7
	42	29.6	11.5	1313	2	US-08-356-060A-4
	43	29.6	11.5	1313	3	US-08-460-900C-4
	44	29.6	11.5	1489	3	US-08-836-582-1
	45	29.6	11.5	2485	1	US-08-424-424B-1

ALIGNMENTS

RESULT 1  
US-08-826-246-9  
; Sequence 9, Application US/08826246  
; Patent No. 6048709  
; GENERAL INFORMATION:  
; APPLICANT: Falb, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/826,246  
; FILING DATE: 28-MAR-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/799,910  
; FILING DATE: 13-FEB-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/011,787  
; FILING DATE: 16-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-078-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)7909090  
; TELEFAX: (212)8699741  
; TELEX: 66141-PENNIE  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1228 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 1...468  
; OTHER INFORMATION:

```
unclassified
REFERENCE 1 (bases 1 to 297)
AUTHORS  Sibson,D.R. and Hadfield,K.M.
TITLE     HUMAN NUCLEIC ACID FRAGMENTS ISOLATED FROM BRAIN, ADRENAL TISSUE,
          PLACENTA OR BONE MARROW AND THEIR USE
JOURNAL   Patent: EP 0587279-A 16-MAR-1994;
          MEDICAL RES COUNCIL (GB)
FEATURES  Location/Qualifiers
           source
             1..297
               /organism="unidentified"
               /db_xref="taxon:32644"
BASE COUNT 66 a 117 c 61 g 52 t 1 others
ORIGIN

Query Match      17.1%; Score 44; DB 81; Length 297;
Best Local Similarity 77.9%; Pred. NO. 4.3;
Matches 53; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY  2  TCCGGCCGCGCTGCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTCTTTCTGCTGC 61
      |||||
DB  225 TCCGGCCGCGCTGCCAGTTAAGGACCAACCTGCCAGAGGCTTCTTTTCGGGTTGA 284
      |||||

QY  62  TCACCATC 69
      |||||
DB  285 TCGCCATC 292
```

Search completed: January 30, 2001, 19:24:37  
Job time: 25963 sec

124276 124375: gap of unknown length  
124376 126296: contig of 1921 bp in length  
126297 126396: gap of unknown length  
126397 129180: contig of 2784 bp in length  
129181 129280: gap of unknown length  
129281 132785: contig of 3405 bp in length  
132686 132785: gap of unknown length  
132786 134635: contig of 1850 bp in length  
134636 134735: gap of unknown length  
134736 137197: contig of 2462 bp in length  
137198 137297: gap of unknown length  
137298 139569: contig of 2272 bp in length  
139570 139669: gap of unknown length  
139670 142014: contig of 2345 bp in length  
142015 142114: gap of unknown length  
142115 144074: contig of 1960 bp in length  
144075 144174: gap of unknown length  
144175 146482: contig of 2308 bp in length  
146483 146582: gap of unknown length  
146583 148126: contig of 1544 bp in length  
148127 148226: gap of unknown length  
148227 150342: contig of 2116 bp in length  
150343 150442: gap of unknown length  
150443 151791: contig of 1349 bp in length  
151792 151891: gap of unknown length  
151892 153769: contig of 1878 bp in length  
153770 153869: gap of unknown length  
153870 156824: contig of 2955 bp in length  
156825 156924: gap of unknown length  
156925 158316: contig of 1392 bp in length  
158317 158416: gap of unknown length  
158417 160019: contig of 1603 bp in length  
160020 160119: gap of unknown length  
160120 161830: contig of 1711 bp in length  
161831 161930: gap of unknown length  
161931 163680: contig of 1750 bp in length  
163681 163780: gap of unknown length  
163781 165772: contig of 1992 bp in length  
165773 165872: gap of unknown length  
165873 167356: contig of 1484 bp in length  
167357 167456: gap of unknown length  
167457 169605: contig of 2149 bp in length  
169606 169705: gap of unknown length  
169706 171781: contig of 2076 bp in length  
171782 171881: gap of unknown length  
171882 173350: contig of 1469 bp in length  
173351 173450: gap of unknown length  
173451 175539: contig of 2089 bp in length  
175540 175639: gap of unknown length  
175640 176975: contig of 1336 bp in length  
176976 177075: gap of unknown length  
177076 178200: contig of 1445 bp in length  
178201 178620: gap of unknown length  
178621 179634: contig of 1014 bp in length  
179635 179734: gap of unknown length  
179735 181530: contig of 1796 bp in length  
181531 181630: gap of unknown length  
181631 182891: contig of 1261 bp in length  
182892 182991: gap of unknown length  
182992 184207: contig of 1216 bp in length  
184208 184307: gap of unknown length  
184308 185488: contig of 1181 bp in length  
185489 185588: gap of unknown length  
185589 186739: contig of 1151 bp in length  
186740 186839: gap of unknown length  
186840 187868: contig of 1029 bp in length.

FEATURES  
Source  
1..187868  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-329K24"

BASE COUNT 46889 a 45196 c 44010 g 45923 t 5850 others

ORIGIN

Query Match 54.0%; Score 139.2; DB 62; Length 187868;  
Best Local Similarity 74.8%; Pred. No. 2e-21;  
Matches 193; Conservative 0; Mismatches 53; Indels 12; Gaps 1;

QY 13 CTGCCAGTCGAGAACCCAGCCAAAGGCTTCTTTCTGCTGCTCACCATCGTC 72  
DB 39433 CTACCAACCGAGAACCCCAACATTGCCAAGAGGTCCTTTCTCTGTTCGCATCATC 39492

QY 73 TTTCGCAGATCCTGATGCTGAAGAGGTTGCCCGGCCCTGCTCCAGAGGACGCC 132  
DB 39493 TTTCGCAGATTTGATGCTGAAGAGGTTGTGCCAGCCCTGCTCCGAGGATGCT 39552

QY 133 CCTAACGCCGCATCC-----TGGCGCCACACCCCTGTGTGCCCGTCTCGAG 180  
DB 39553 ACCAGCGCGTGACACCTGAGCCCATTTCTGCGCCCATTTACTGCGCCCTCGAG 39612

QY 181 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCGACCTCAGCATTTCCTCCAG 240  
DB 39613 CTTTGAACCTGACCTCGGAGTCTCGACTATGCGCTGATGCTTAAGCTTTCTCCAG 39672

QY 241 CAACACCGCGCGCCTTC 258  
DB 39673 CAACATCCGCGCGCCTTC 39690

RESULT 14  
A75448

LOCUS A75448 297 bp DNA PAT 15-OCT-1999  
DEFINITION Sequence 1134 from Patent WO9401548.  
ACCESSION A75448  
VERSION A75448.1 GI:6065525  
SOURCE  
ORGANISM  
unidentified.  
unclassified.  
REFERENCE 1 (bases 1 to 297)  
AUTHORS Sibson,D.R. and Gross,J.  
TITLE HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN ADRENAL TISSUE,  
PLACENTA OR BONE MARROW  
JOURNAL Patent: WO 9401548-A 20-JAN-1994;  
MEDICAL RES COUNCIL (GB); SIBSON DAVID ROSS (GB)  
FEATURES  
source  
1..297  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 66 a 117 c 61 g 52 t 1 others

ORIGIN

Query Match 17.1%; Score 44; DB 81; Length 297;  
Best Local Similarity 77.9%; Pred. No. 4.3;  
Matches 53; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2 TCCGGCGCCAGCTGCCAGTCGAGGAACCCAGCCACCCAAAGGCTTCTTTCTCTCGTC 61  
DB 225 TCCGGCGCAGTTGCCAGTTAAGGACCAACCTCGCCAGAGGCTTCTTTTCGGGTGA 284

QY 62 TCACCATC 69  
DB 285 TCGCCATC 292

RESULT 15  
A78427

LOCUS A78427 297 bp DNA PAT 19-OCT-1999  
DEFINITION Sequence 1134 from Patent EP0587279.  
ACCESSION A78427  
VERSION A78427.1 GI:6090092  
KEYWORDS  
SOURCE  
ORGANISM  
unidentified.  
unclassified.

```
/strain="BALB/c"
/db_xref="taxon:10090"
BASE COUNT 390 a 525 c 503 g 520 t
ORIGIN

Query Match 58.0%; Score 149.6; DB 11; Length 1938;
Best Local Similarity 75.6%; Pred. No. 2.7e-23;
Matches 204; Conservative 0; Mismatches 54; Indels 12; Gaps 1;

OY 1 GTCCGGCCGAGTCCAGTCGAGGAGACCCAGCCAAAGAGGCTTCTTTCGCTC 60
DB 1031 GTCCGGCCGAGTACCAACGAGGAGACCCAAACATTTGCCAAGAGGCTCTTTCCTG 1090

OY 61 CTCACCATCTCTTCTGCCAGATCCTGATGCTGAAGAGGCTGCGCGGCCCTCCCTCC 120
DB 1091 TTCCCATCATCTTCTGCCAGATTTGATGGCTGAAGAGGCTGTGCGAGCCCCCTGGCT 1150

OY 121 CCAGAGAGCGCCCTAACGGCGCATCC-----TGCGCGCCAGCCCTGTGTGTC 168
DB 1151 CCGAGGATGCTACCAAGCGCGTGACACCTTGAGCCCATTTCTCGCGCCATTAATCGGCC 1210

OY 169 CCCTCTCCAGCCCTTTATCTGACTTCGAGCTCGAGCCCTCGGACTACGCTCGGACCTCAGC 228
DB 1211 CCGCTCTCCAGCCCTTTGAACCTTGACCTCGGAGTCTCGGACTGATGCTGGATCTTAA 1270

OY 229 ACTTTCCTCCAGCAACACCGCGCGCTTC 258
DB 1271 GCTTTCTCTAGCAACATCGCGCGCTTC 1300

RESULT 13
AC022301 187868 bp DNA HTG 15-JUN-2000
LOCUS Mus musculus clone Rp23-329K24, *** SEQUENCING IN PROGRESS ***, 59
DEFINITION unordered pieces.
ACCESSION AC022301.9 GI:8493517
VERSION HTG; HTGS_PHASE1.
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Pederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Loulsegod,H., Lozado,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogihara,M., Parish,B.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williams,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstein,G.,
Worley,K. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 187868)
Worley,K.C.
Direct Submission
Submitted (29-JAN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 13, 2000 this sequence version replaced gi:8248590.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
```



COMMENT

-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----

Project Information  
Center Project Name: 0  
Center clone name: RPCI-23\_128D3  
-----

Summary Statistics  
Consensus quality: 139878 bases at least Q40  
Consensus quality: 165603 bases at least Q30  
Consensus quality: 177139 bases at least Q20  
Estimated insert size: 183300; agarose-fp estimation  
Estimated insert size: 200576; sum-of-contigs estimation  
Quality coverage: 3.48 in Q20 bases; agarose-fp estimation  
Quality coverage: 3.18 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 30 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1175: contig of 1175 bp in length  
1176: gap of unknown length  
1276: contig of 1288 bp in length  
2563: gap of unknown length  
2663: contig of 1243 bp in length  
3906: gap of unknown length  
4006: contig of 1218 bp in length  
5224: gap of unknown length  
5324: contig of 1312 bp in length  
6636: gap of unknown length  
6736: contig of 2460 bp in length  
9196: gap of unknown length  
9296: contig of 1427 bp in length  
10223: gap of unknown length  
10823: contig of 2029 bp in length  
12852: gap of unknown length  
12953: contig of 3073 bp in length  
16026: gap of unknown length  
16126: contig of 2971 bp in length  
19096: gap of unknown length  
19196: contig of 2554 bp in length  
21750: gap of unknown length  
21850: contig of 4729 bp in length  
26579: gap of unknown length  
26680: contig of 5793 bp in length  
32473: gap of unknown length  
32573: contig of 4422 bp in length  
36994: gap of unknown length  
37094: contig of 6977 bp in length  
44071: gap of unknown length  
44171: contig of 6556 bp in length  
50727: gap of unknown length  
50827: contig of 6377 bp in length  
57204: gap of unknown length  
57304: contig of 7772 bp in length  
65076: gap of unknown length  
65176: contig of 6648 bp in length  
71824: gap of unknown length  
71924: contig of 6111 bp in length  
78033: gap of unknown length  
78133: contig of 9703 bp in length  
87838: gap of unknown length  
87839: contig of 6986 bp in length  
94924: gap of unknown length  
95024: contig of 6164 bp in length  
95025: gap of unknown length  
101188: contig of 7637 bp in length  
101289: gap of unknown length  
108925: contig of 1175 bp in length  
108926: gap of unknown length

\* 109026 119738: contig of 10713 bp in length  
\* 119739 119838: gap of unknown length  
\* 119839 129876: contig of 10038 bp in length  
\* 129877 129976: gap of unknown length  
\* 129977 143243: contig of 13267 bp in length  
\* 143244 161164: gap of unknown length  
\* 143344 161264: contig of 17821 bp in length  
\* 161165 161264: gap of unknown length  
\* 161265 180316: contig of 19052 bp in length  
\* 180317 180416: gap of unknown length  
\* 180417 203476: contig of 23060 bp in length.  
FEATURES  
Location/Qualifiers  
1..203476  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-128D3"  
/clone\_lib="RPCI mouse BAC library 23"  
BASE COUNT 51748 a 48714 c 48138 g 51580 t 3296 others  
ORIGIN

Query Match 58.6%; Score 151.2; DB 70; Length 203476;  
Best Local Similarity 75.9%; Pred. No. 4.7e-24;  
Matches 205; Conservative 0; Mismatches 53; Indels 12; Gaps 1;  
QY 1 GTCGGGGCCAGCTGCCAGTCGAGGAACCCAGCAACCCAGCAAAAGGCTTCTCTTTCTGCTG 60  
|||||  
DB 92875 GTCGGGGCCAGCTACCAACCGGAGGAGCAACCAATGCGAAGAGGCTCTCTTTCTCCTG 92934  
QY 61 CTCACCATCGCTTCTGCCAGATCCTGATGGTGAAGAGGCTGTCGGCGGCCCTGCT 120  
|||||  
DB 92935 TTCGCCATCATCTTCTGCCAGATTTTGTGCTGAAGAGGCTGTCGACGCCCTGGCT 92994  
QY 121 CCAGAGGAGCCCTTAACGGCGCATCCC-----TGGCGCCACCCCTGTGTC 168  
DB 92995 CCGGAGGATGCTACGAGCGCTGACACCTGAGCCATTTCTGGCCCATTTACTGCGCCC 93054  
QY 169 CCCGTCTCGAGCCCTTAATCTGACTTCGGAGCCCTCGGACTAGGCTCTGGACCTCAGC 228  
|||||  
DB 93055 CCGTCTCGAGCCCTTGAACCTGACCTCGAGTCTCGACTATGCGCTGGTCTTAA 93114  
QY 229 ACTTCTCCAGCAACCCCGCCCTTC 258  
|||||  
DB 93115 GCTTCTCCAGCAACATCCGGCGCCTTC 93144

RESULT 12  
MMGLY96 1938 bp mRNA ROD 11-MAR-1993  
LOCUS M.musculus gly96 mRNA.  
DEFINITION X67644  
ACCESSION X67644.1 GI:287803  
VERSION gly96 gene; glycosylated protein.  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1938)  
AUTHORS Lau, L.  
TITLE Direct Submission  
JOURNAL Submitted (29-JUL-1992) L. Lau, University of Illinois College of  
Med., Dept. of Genetics, M/C 669, 808 South Wood Street, Chicago,  
IL 60612, USA  
REFERENCE 2 (bases 1 to 1938)  
AUTHORS Charles, C.H., Yoon, J.K., Simske, J.S. and Lau, L.F.  
TITLE Genomic structure; cDNA sequence, and expression of gly96, a growth  
factor-inducible immediate-early gene encoding a short-lived  
glycosylated protein  
JOURNAL Oncogene 8 (3), 797-801 (1993)  
MEDLINE 93173526  
FEATURES Location/Qualifiers  
1..1938  
/organism="Mus musculus"



```

Db 61000 GTCCGGCGCAGCTGCAGTTCGAGACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 61059
Qy 61 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTTCCT 120
Db 61060 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTTCCT 61119
Qy 121 CCAGAGACGCCCTTAACGGCGGATCCTGTGGCGCCACCCCTGTGTGCCCGTCTCTCGAG 180
Db 61120 CCAGAGACGCCCTTAACGGCGGATCCTGTGGCGCCACCCCTGTGTGCCCGTCTCTCGAG 61179
Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACTTCTCCCTCCAG 240
Db 61180 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACTTCTCCCTCCAG 61239
Qy 241 CAACACCCGCCCTTC 258
Db 61240 CAACACCCGCCCTTC 61257

RESULT 8
AP000512 200000 bp DNA PRI 30-MAR-2000
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region,
DEFINITION section 11/20.
ACCESSION AP000512
VERSION AP000512.1 GI:5926699
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Shiina,S., Tamliya,G., Oka,A. and Inoko,H.
Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region
Published Only in DataBase (1999) In press
2 (bases 1 to 200000)
Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
Direct Submission
Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Mika
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail:mika@tokyo.jst.go.jp,
URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)
This sequence is conducted by Tokai University as a JST sequencing
Team.
Principal Investigator: Hidetoshi Inoko Ph.D
Phone:+81-463-93-1121, Fax:+81-463-94-8884,
The sequence is submitted by Human Genome Sequencing in ALIS
project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences, please visit our
sequence archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/top.
html) or send email to webmaster@www-alis.tokyo.jst.go.jp.
FEATURES
source
1..200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
101440..101648
/feature="stSG4264:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4566033"
complement(114197..114328)
/standard_name="D6S2115"
/feature="SHGC-13451:The location is between each flanking
site of PCR primers."
/db_xref="GDB:738998"
complement(114300..114430)
/feature="WI-18721:The location is between each flanking site

```

```

of PCR primers."
/db_xref="GDB:4574312"
117236..117362
/feature="SGC34396:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4580073"
complement(117712..117836)
/standard_name="D6S2041"
/feature="SHGC-10364:The location is between each flanking
site of PCR primers."
/db_xref="GDB:735115"
complement(117961..118093)
/feature="SGC32809:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4584619"
142184..142339
/feature="K1A0170:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4586652"
complement(142215..142343)
/feature="WI-15445:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4581086"
complement(153977..154114)
/feature="SGC34551:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4582352"
complement(165191..165295)
/standard_name="D6S1376"
/feature="downstream/upstream:The location is between each
flanking site of PCR primers."
/db_xref="GDB:580497"
188925..188989
/feature="pml216:The location is between each flanking site
of PCR primers."
/db_xref="GDB:594325"
complement(197650..197804)
/feature="Cda01e04:The location is between each flanking site
of PCR primers."
/db_xref="GDB:443319"
52458 a 48336 c 49153 g 50053 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 258; DB 51; Length 200000;
Best Local Similarity 100.0%; Pred. No. 2.4e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCCGGCGCAGCTGCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60
Db 97985 GTCCGGCGCAGCTGCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 98044
Qy 61 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTTCCT 120
Db 98045 CTCACCATCTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTTCCT 98104
Qy 121 CCAGAGACGCCCTTAACGGCGGATCCTGTGGCGCCACCCCTGTGTGCCCGTCTCTCGAG 180
Db 98105 CCAGAGACGCCCTTAACGGCGGATCCTGTGGCGCCACCCCTGTGTGCCCGTCTCTCGAG 98164
Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACTTCTCCCTCCAG 240
Db 98165 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGACCTCAGCACTTCTCCCTCCAG 98224
Qy 241 CAACACCCGCCCTTC 258
Db 98225 CAACACCCGCCCTTC 98242

RESULT 9
S81914 1223 bp mRNA PRI 12-AUG-1996
LOCUS IEX-1-radiation-inducible immediate-early gene (human, placenta,
DEFINITION

```

```
/rpt_family="L1"
13869. .14279
/rpt_family="Alu"
18033. .18284
/rpt_family="Alu"
19178. .19451
/rpt_family="Alu"
19458. .19749
/rpt_family="Alu"
20862. .21002
/rpt_family="MERS"
21586. .21832
/rpt_family="Alu"
22231. .22521
/rpt_family="Alu"
22679)
complement(22616. .22679)
/rpt_family="Alu"
23145)
complement(22869. .23145)
/rpt_family="Alu"
23772)
complement(23478. .23772)
/rpt_family="Alu"
24184. .24372
/rpt_family="Alu"
24529. .24813
/rpt_family="Alu"
25342)
complement(25060. .25342)
/rpt_family="Alu"
25653)
complement(25367. .25653)
/rpt_family="Alu"
25787. .26076
/rpt_family="Alu"
26491. .26762
/rpt_family="Alu"
26866. .27143
/rpt_family="Alu"
27919)
complement(27620. .27919)
/rpt_family="Alu"
29525. .29719
/rpt_family="MIR"
29762. .30333
/rpt_family="Alu"
30365. .30653
/rpt_family="Alu"
31430. .31722
/rpt_family="Alu"
31765. .31857
/rpt_family="Alu"
32265. .32563
/rpt_family="Alu"
32858)
complement(32710. .32858)
/rpt_family="Alu"
33584)
complement(33374. .33584)
/rpt_family="MER2"
34595)
complement(34547. .34595)
/rpt_family="MIR"
36954)
complement(36673. .36954)
/rpt_family="Alu"
37230)
complement(37002. .37230)
/rpt_family="Alu"
40170)
complement(40018. .40170)
/rpt_family="Alu"
40452
/nt="high quality variation versus 3' overlapping clone"
40509. .41109
/rpt_family="Alu"
42609. .42892
/rpt_family="Alu"
43240. .43370
/rpt_family="MIR"
43493. .43779
/rpt_family="Alu"
BASE COUNT 11505 a 10560 c 11226 g 10827 t
ORIGIN
```

```
Query Match 100.0%; Score 258; DB 9; Length 44118;
Best Local Similarity 100.0%; Pred. No. 3.3e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCCGGCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTCTTTCTGTG 60
Db 14923 GTCCGGCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTCTTTCTGTG 14982

Qy 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGCCCTCGCT 120
Db 14983 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGCCCTCGCT 15042

Qy 121 CCAGAGGACGCCCTTAACCGCGCATCCCTGGCGCCACCCCTGTCTCCCGCTCTCCAG 180
Db 15043 CCAGAGGACGCCCTTAACCGCGCATCCCTGGCGCCACCCCTGTCTCCCGCTCTCCAG 15102

Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCCAG 240
Db 15103 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCCAG 15162

Qy 241 CAACACCGCGCGCCTTC 258
Db 15163 CAACACCGCGCGCCTTC 15180

RESULT 7
AB023051 90244 bp DNA PRI 20-NOV-1999
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,
DEFINITION clone:876L4, complete sequence.
ACCESSION AB023051
VERSION AB023051.1 GI:5672606
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:876L4.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Shiina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y.,
Watanabe,K., Kuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Bahram,S. and Inoko,H.
TITLE Molecular dynamics of MHC genesis unraveled by sequence analysis of
the 1,796,938-bp HLA class I region
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
MEDLINE 20027539
REFERENCE 2 (bases 1 to 90244)
AUTHORS Shiina,T. and Takishima,N.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1999) to the DDBJ/EMBL/GenBank databases. Takashi
Shiina, Tokai University School of Medicine, Department of
Molecular Life Science 2; Bohseidai, Isehara, Kanagawa 259-1193,
Japan (E-mail:tshiina@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121,
Fax:81-463-94-8884)
FEATURES
Location/Qualifiers
Source 1..90244
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="876L4"
/map="6p21.3"

BASE COUNT 24562 a 21396 c 21572 g 22714 t
ORIGIN

Query Match 100.0%; Score 258; DB 8; Length 90244;
Best Local Similarity 100.0%; Pred. No. 2.8e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCCGGCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTCTTTCTGTG 60
```

Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCCAGCTGCGAGTCCAGTGGAGAACCCAGCCAAAGGCTTCTTTTCGTG 60  
Db 914 GTCCGGCCAGCTGCGAGTCCAGTGGAGAACCCAGCCAAAGGCTTCTTTTCGTG 973  
QY 61 CTCACCATCGTCTTCTCCAGATCCCTGATGGCTGAAGAGGTGTGCCGGCCCTTCCCT 120  
Db 974 CTCACCATCGTCTTCTCCAGATCCCTGATGGCTGAAGAGGTGTGCCGGCCCTTCCCT 1033  
QY 121 CCAGAGAGCCCTTAACGCCGATCCCTGGCGCCACCCCTGTGTGCCCGCTCCCTCGAG 180  
Db 1034 CCAGAGAGCCCTTAACGCCGATCCCTGGCGCCACCCCTGTGTGCCCGCTCCCTCGAG 1093  
QY 181 CCCTTTAATCTGACTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCCCGAG 240  
Db 1094 CCCTTTAATCTGACTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCCCGAG 1153  
QY 241 CAACACCCGCCGCTTC 258  
Db 1154 CAACACCCGCCGCTTC 1171

RESULT 6

AC006165 44118 bp DNA PRI 08-DEC-1998  
LOCUS Homo sapiens clone UWGC:y54c125 from 6p21, complete sequence.  
DEFINITION AC006165  
ACCESSION AC006165  
VERSION AC006165.1 GI:3980464  
KEYWORDS HTG.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 44118)  
AUTHORS Janer, M., Guillaudoux, T., Vu, Q., Kutayavin, T., Harter, H. and Geraghty, D. E.

TITLE Large scale sequence analysis of the human MHC class I region

JOURNAL Unpublished (1998)

REMARK Fred Hutchinson Cancer Research Center  
The Clinical Research Division  
1100 Fairview Ave. N., P.O. Box 19024  
Seattle, WA 98109-1024

2 (bases 1 to 44118)  
Geraghty, D. E. and Olson, M. V.

Direct Submission  
Submitted (08-DEC-1998) Human Genome Center, University of  
Washington, Box 352145, Seattle, WA 98195, USA

University of Washington Human Genome Center  
Box 352145, Seattle, WA 98195

Contact: Daniel E. Geraghty (geraghty@fhcrc.org)

Overlapping Sequences:  
5': UWGC:y54c283  
3': UWGC:y54c222 (Genbank Accession: AC006049)

Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.

Double stranded (DS) coverage: 58.8%

DS or two chemistry coverage: 99.5%

Single stranded regions: 2

Sequence Validation:  
This sequence has been validated by Multiple Complete Digest  
Mapping. Comparison of the experimentally derived map digest  
fragments with sequence-predicted fragments is given below.

Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

BgIII		HindIII		NsiI	
Map	Seq	Map	Seq	Map	Seq
6714.50	6452.00	1823.23	1794.00	5581.12	5401.00
1089.04	1085.00	1246.04	1228.00	10570.24	10123.00
1536.15	1528.00	5570.33	5372.00	10069.29	9626.00
3623.08	3537.00	1670.92	1630.00		
3784.08	3621.00	1270.76	1280.00		
6834.00	6689.00	4235.79	4109.00		
11234.36	10874.00	1171.00	1151.00		
762.43	762.00	3226.11	3096.00		
2727.38	2653.00	10501.64	10080.00		
1691.42	1684.00	4208.44	4052.00		
1295.60	1277.00				
746.87	734.00				
FEATURES					
Source					
1. 44118					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/chromosome="6"					
/map="6p21"					
/clone="CGM1:D100E8"					
/sub_clone="UWGC:y54c125"					
/cell_line="CGM1"					
/clone_lib="Wash U YAC library"					
218..405					
/rpt_family="MER3"					
400..999					
/rpt_family="Alu"					
1797..2378					
/rpt_family="Alu"					
2649..2956					
/rpt_family="Alu"					
complement(3372..3894)					
/rpt_family="Alu"					
3941..4070					
/rpt_family="Alu"					
complement(449..4724)					
/rpt_family="Alu"					
complement(5117..5389)					
/rpt_family="Alu"					
7668..7913					
/rpt_family="Alu"					
complement(7924..8226)					
/rpt_family="Alu"					
8250..8551					
/rpt_family="Alu"					
complement(8893..8973)					
/rpt_family="MIR"					
complement(11829..12096)					
/rpt_family="Alu"					
12416..12740					
/rpt_family="Alu"					
complement(13181..13247)					
/rpt_family="Alu"					
complement(13498..13647)					

RESULT 4

AF071596 1693 bp DNA PRI 10-NOV-1998

LOCUS Homo sapiens apoptosis inhibitor (IEX-1L) gene, complete cds.

DEFINITION AF071596

ACCESSION AF071596

VERSION AF071596.1 GI:3851531

KEYWORDS human.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1693)

AUTHORS Wu.M.X., Ao.Z., Prasad.K.V., Wu.R. and Schlossman.S.F.

TITLE IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival

JOURNAL Science 281 (5379), 998-1001 (1998)

MEDLINE 98369175

REFERENCE 2 (bases 1 to 1693)

AUTHORS Wu.M.X. and Ao.Z.

TITLE Direct Submission

JOURNAL Submitted (10-JUN-1998) Tumor Immunology, Dana Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA

FEATURES

source

1..1693

/organism="Homo sapiens"

/db\_xref="taxon:9606"

359..1693

/gene="IEX-1L"

/product="apoptosis inhibitor"

359..1693

/gene="IEX-1L"

385..966

/gene="IEX-1L"

/note="immediate early response protein"

/codon\_start=1

/product="apoptosis inhibitor"

/protein\_id="AAC72344.1"

/db\_xref="GI:3851532"

/translation="MCHSRSCPTMTILOAPTAPSTIPGPRKSGPEITFDPLPEP AAAPAGPSASRGRKRRVLPVPRVVRQLPVEEPNPKRLI.FULLTIVFCOII.MAE EGVDPALPPEDAPNAASLAPTVPSPVLEPNTSEPSDYALDLSTFIQQHPAAF"

BASE COUNT 348 a 529 c 456 g 360 t

ORIGIN

Query Match 100.0%; Score 258; DB 10; Length 1693;

Best Local Similarity 100.0%; Pred. No. 6.4e-47;

Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCCGGCCAGCTCCAGTCGAGGACCGAACCCAGCCAGCAAGGCTTCTCTTCTGCTG 60

DB 706 GTCCGGCCAGCTCCAGTCGAGGACCGAACCCAGCCAGCAAGGCTTCTCTTCTGCTG 765

OY 61 CTCACCATCGTCTTCTGCGAGATCGTGTGCTGAAGAGGGTGTGCCGGCCGCCCTGCCT 120

DB 766 CTCACCATCGTCTTCTGCGAGATCGTGTGCTGAAGAGGGTGTGCCGGCCGCCCTGCCT 825

OY 121 CCAGAGGAGCCCTTAAGCCGCCATCCCTGCGCCGCCACCCCTGCTGCTCCCGCTCCAG 180

DB 826 CCAGAGGAGCCCTTAAGCCGCCATCCCTGCGCCGCCACCCCTGCTGCTCCCGCTCCAG 885

OY 181 CCCTTAATCTGACTTCGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCTCCAG 240

DB 886 CCCTTAATCTGACTTCGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCTCCAG 945

OY 241 CAACACCCGGCCCTTC 258

DB 946 CAACACCCGGCCCTTC 963

RESULT 5

HSPRG1

LOCUS 1864 bp DNA PRI 06-NOV-1998

DEFINITION H.sapiens PRG1 gene.

ACCESSION X96438

VERSION X96438.1 GI:2440072

KEYWORDS PRG1 gene.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1864)

AUTHORS Trauzold.A.

TITLE Direct Submission

JOURNAL Submitted (05-MAR-1996) A. Trauzold, Laboratory of Molecular Gastroenterology, 1st Dept. of Medicine, University of Kiel, Schittenhelmstrasse 12, Kiel, D-24105, FRG

REMARK revised by [4]

REFERENCE 2 (bases 1 to 508)

AUTHORS Schaefer.H., Trauzold.A., Lettau.P., Kalthoff.H., Foelsch,U.R. and Schmidt,W.E.

TITLE cDNA cloning and sequencing of a novel human early response gene and characterization of its expression in pancreatic carcinoma cells

JOURNAL Gastroenterology In press

REFERENCE 3 (bases 1 to 1864)

AUTHORS Schaefer.H.

TITLE Direct Submission

JOURNAL Submitted (23-SEP-1997) H.Schaefer, Trauzold, Laboratory of Molecular Gastroenterology, 1st Dept. of Medicine, University of Kiel, Schittenhelmstrasse 12, Kiel, D-24105, FRG

REFERENCE 4 (bases 1 to 1864)

AUTHORS Schaefer.H., Diebel.J., Arlt.A., Trauzold.A. and Schmidt.W.E.

TITLE The promoter of human p22/PACAP response gene 1 (PRG1) contains functional binding sites for the p53 tumor suppressor and for NFkappaB

JOURNAL FEBS Lett. 436 (2), 139-143 (1998)

MEDLINE 98452926

COMMENT On Sep 27, 1997 this sequence version replaced gi:1515291.

FEATURES

source

1..1864

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_line="818-4"

1..562

/gene="PRG1"

1..1864

/gene="PRG1"

Join(563..801,914..1864)

/gene="PRG1"

563..801

/gene="PRG1"

/number=1

Join(592..801,914..1174)

/gene="PRG1"

/codon\_start=1

/protein\_id="CAA65304.1"

/db\_xref="GI:2440073"

/db\_xref="SWISS-PROT:P46695"

/translation="MCHSRSCPTMTILOAPTAPSTIPGPRKSGPEITFDPLPEP AAAPAGPSASRGRKRRVLPVPRVVRQLPVEEPNPKRLI.FULLTIVFCOII.MAE EGVDPALPPEDAPNAASLAPTVPSPVLEPNTSEPSDYALDLSTFIQQHPAAF"

802..913

/gene="PRG1"

/number=1

914..1864

/gene="PRG1"

/number=2

BASE COUNT 395 a 571 c 519 g 379 t

ORIGIN

Query Match 100.0%; Score 258; DB 53; Length 1864;

Best Local Similarity 100.0%; Pred. No. 6.3e-47;

```
/db_xref="taxon:9606"
/cell_type="leukocytes"
<!.il
/gene="dif-2"
1..1230
/gene="dif-2"
12..482
/gene="dif-2"
/codon_start=1
/product="Dif-2 protein"
/protein_id="CAA74886.1"
/db_xref="GI:2660551"
/db_xref="SWISS-PROT:P46695"
/translation="MCHSRCHPTMTILOAPTTPABSTIPGPRGSGPELFTFDPLPEP
AAPAGRSASRGHRKRRVLYPRVRRQLPVEEPNPAKRLFLLLLTIVFCQILMAE
EGVPAPLPEDAPNAASLAPTPVSPVLEPFLNTSEPSYALDLSFLQQHPAAF"
172
conflict
/gene="dif-2"
/citation={3}
/replace="g"
328
/gene="dif-2"
/citation={3}
/replace="g"
483..>1230
/gene="dif-2"
673..675
/gene="dif-2"
/citation={3}
/replace="ag"
687..689
/gene="dif-2"
/citation={3}
/replace="tg"
783..785
/gene="dif-2"
/citation={3}
/replace="tg"
1020..1021
/gene="dif-2"
/citation={3}
/replace="att"
1120
/gene="dif-2"
/citation={3}
/replace="g"
1133
/gene="dif-2"
/citation={3}
/replace="t"
BASE COUNT 259 a 367 c 342 g 262 t
ORIGIN

Query Match 100.0%; Score 258; DB 53; Length 1230;
Best Local Similarity 100.0%; Pred. No. 6.8e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGCTGCCAGTCAGGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60
DB 222 GTCCGGCGCCAGCTGCCAGTCAGGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 281

QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGGCT 120
DB 282 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGGCT 341

QY 121 CCAGAGGAGCGCCCTAACCGCCATCCCTGGCGCCACCCCTGCTCCCGCTCCGAG 180
DB 342 CCAGAGGAGCGCCCTAACCGCCATCCCTGGCGCCACCCCTGCTCCCGCTCCGAG 401

QY 181 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCAGCACTTTCTCCAG 240
DB 402 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTCGACCTCAGCACTTTCTCCAG 461
```

```
QY 241 CAACACCCGGCGCCCTTC 258
DB 462 CAACACCCGGCGCCCTTC 479

RESULT 3
AF039067 1309 bp mRNA PRI 21-AUG-1998
LOCUS Homo sapiens anti-death protein (IEX-1L) mRNA, complete cds.
DEFINITION AF039067
ACCESSION AF039067
VERSION AF039067.1 GI:3449375
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1309)
AUTHORS Wu,M.X., Ao,Z., Prasad,K.V.S., Wu,R. and Schlossman,S.F.
TITLE IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell
survival
JOURNAL Science 281 (5379), 998-1001 (1998)
MEDLINE 98369175
REFERENCE 2 (bases 1 to 1309)
AUTHORS Ao,Z. and Wu,M.X.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1997) Tumor Immunology, Dana Farber Cancer
Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES
Source
1..1309
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1309
/gene="IEX-1L"
1..582
/gene="IEX-1L"
/note="type II membrane protein"
/codon_start=1
/product="anti-death protein"
/protein_id="AAC32558.1"
/db_xref="GI:3449376"
/translation="MCHSRCHPTMTILOAPTTPABSTIPGPRGSGPELFTFDPLPEP
AAPAGRSASRGHRKRRVLYPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAE
EGVPAPLPEDAPNAASLAPTPVSPVLEPFLNTSEPSYALDLSFLQQHPAAF"
BASE COUNT 262 a 393 c 369 g 285 t
ORIGIN

Query Match 100.0%; Score 258; DB 10; Length 1309;
Best Local Similarity 100.0%; Pred. No. 6.7e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGCTGCCAGTCAGGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60
DB 322 GTCCGGCGCCAGCTGCCAGTCAGGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTC 381

QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGGCT 120
DB 382 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCCGGCGCCCTGGCT 441

QY 121 CCAGAGGAGCGCCCTAACCGCCATCCCTGGCGCCACCCCTGCTCCCGCTCCGAG 180
DB 442 CCAGAGGAGCGCCCTAACCGCCATCCCTGGCGCCACCCCTGCTCCCGCTCCGAG 501

QY 181 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCAGCACTTTCTCCAG 240
DB 502 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGAGCTCAGCACTTTCTCCAG 561

QY 241 CAACACCCGGCGCCCTTC 258
DB 562 CAACACCCGGCGCCCTTC 579
```

```

c 22 39.4 15.3 188907 65 AC026986
c 23 39.2 15.2 150152 74 AL355502
c 24 39 15.1 144798 6 AC026758
c 25 38.4 14.9 10165 1 AE002089
c 26 38.2 14.8 4391 52 HSAJ4832
c 27 38.2 14.8 8391 29 MBU36763
c 28 38.2 14.8 38390 29 SC2H12
c 29 38.2 14.8 174044 69 AC069521
c 30 38 14.7 580 11 MMU80602
c 31 38 14.7 64990 51 AP000251
c 32 38 14.7 100000 34 AP000030
c 33 38 14.7 100000 34 AP000134
c 34 38 14.7 100000 51 AP000212
c 35 38 14.7 340000 51 AP001711
c 36 37.8 14.7 135119 56 AC011578
c 37 37.6 14.6 162922 70 AC074092
c 38 37.6 14.6 196978 58 AC015872
c 39 37.4 14.5 65671 71 AC083783
c 40 37.2 14.4 843 77 CNS01K6M
c 41 37.2 14.4 2043 11 AF102818
c 42 37.2 14.4 12664 1 AF005029
c 43 37.2 14.4 50502 52 HSA494016
c 44 37.2 14.4 69354 69 AC069342
c 45 37 14.3 576 77 CNS01J02

```

## ALIGNMENTS

```

RESULT 1
AF083421 477 bp mRNA PRI 02-SEP-1998
LOCUS Homo sapiens radiation-inducible immediate early response gene IEX1
DEFINITION (IEX1) mRNA, complete cds.
ACCESSION AF083421
VERSION AF083421.1 GI:3511288
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 477)
Kondratyev,A.D., Chung,K.N. and Jung,M.O.
Identification and characterization of a radiation-inducible
glycosylated human early-response gene
Cancer Res. 56 (7), 1498-1502 (1996)
JOURNAL 96181295
MEDLINE 2 (bases 1 to 477)
Kumar,R., Kobayashi,T., Warner,G.M., Wu,Y., Salisbury,J.L.,
Lingle,W. and Pittelkow,M.R.
A Novel Immediate Early Response Gene, IEX-1 Is Induced by
Ultraviolet Radiation in Human Keratinocytes
Unpublished
JOURNAL 3 (bases 1 to 477)
Kumar,R., Pittelkow,M.R. and Warner,G.M.
Direct Submission
AUTHORS Submitted (11-AUG-1998) Nephrology Research, Mayo Clinic, 200 1st
TITLE Street SW, Rochester, MN 55905, USA
JOURNAL Location/Qualifiers
FEATURES
source
1. .477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="keratinocyte; fetal osteoblast"
gene 1. .477
/genes="IEX1"
CDS 7. .477
/genes="IEX1"
/feature="UV irradiation induced gene in keratinocytes"
/codon_start=1
/product="radiation-inducible immediate early response
gene IEX1"
/protein_id="AAC33793.1"
/db_xref="GI:3511288"

```

```

/translation="MCHSRSCHPTMTIIQAPTHAPSTIPGPRKSGCPLEIFTFDPLPEP
AAPAGRPASGRHRRKRVLYPRVRQLPVEEPNPAKRLLLILLITVFCQILMAE
EGVPAPLPEDAPNAASLAPTVPSPVLEFNLTSSESDYALDLSTFLQOHPAAF"
167
/genes="IEX1"
/citation=[1]
/replace="g"
323
/genes="IEX1"
/citation=[1]
/replace="g"
BASE COUNT 72 a 205 c 117 g 83 t
ORIGIN
Query Match 100.0%; Score 258; DB 10; Length 477;
Best Local Similarity 100.0%; Pred. No. 8.3e-47; Indels 0; Caps 0;
Matches 258; Conservative 0; Mismatches 0;
Qy 1 GTCCGGCGCCAGCTGCCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 60
Db 217 GTCCGGCGCCAGCTGCCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTCTTCTGCTG 276
Qy 61 CTCACCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGCTGTGCCGGCCCTTCCCT 120
Db 277 CTCACCATCGTCTTCTGCCAGATCCTGATGCTGAAGAGGCTGTGCCGGCCCTTCCCT 336
Qy 121 CCAGAGGAGCCCTTAACGGCGATCCTGGCGCCACCCCTGTGTCCTCCCGTCTCCGAG 180
Db 337 CCAGAGGAGCCCTTAACGGCGATCCTGGCGCCACCCCTGTGTCCTCCCGTCTCCGAG 396
Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACATCCTGACCTCTGGACCTCAGACATTCCTCCAG 240
Db 397 CCCTTTAATCTGACTTCGGAGCCCTCGGACATCCTGACCTCTGGACCTCAGACATTCCTCCAG 456
Qy 241 CAACACCGCGCCGCTTC 258
Db 457 CAACACCGCGCCGCTTC 474
RESULT 2
HSDIF2 1230 bp mRNA PKI 12-AUG-1997
LOCUS Homo sapiens mRNA for DIF-2 protein.
DEFINITION Y14551
ACCESSION Y14551
VERSION Y14551.1 GI:2660550
KEYWORDS dif-2 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1230)
Pietzsch,A., Buchler,C., Aslanidis,C. and Schmitz,G.
Identification and characterization of a novel monocyte/macrophage
differentiation-dependent gene that is responsive to
lipopolysaccharide, ceramide, and lysophosphatidylcholine
Biochem. Biophys. Res. Commun. 235 (1), 4-9 (1997)
JOURNAL 97339426
MEDLINE 2 (bases 1 to 1230)
Pietzsch,A.
AUTHORS Direct Submission
TITLE Submitted (11-AUG-1997) A. Pietzsch, Institute for Clinical
JOURNAL Chemistry and Laboratory Medicine, University of Regensburg,
Franz-Josef-Strauss-Allee 11, 93053 Regensburg, FRG
REFERENCE 3 (bases 1 to 1230)
Kondratyev,A.D., Chung,K.N. and Jung,M.O.
Identification and characterization of a radiation-inducible
glycosylated human early-response gene
Cancer Res. 56 (7), 1498-1502 (1996)
JOURNAL 96181295
MEDLINE Location/Qualifiers
FEATURES
source
1. .1230
/organism="Homo sapiens"

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 19:23:39 ; Search time 1900.13 Seconds  
(without alignments)  
694.889 Million cell updates/sec

Title: US-08-799-910-9\_COPY\_211\_468

Perfect score: 258

Sequence: 1 GTCCGGCCCGCCAGTCCAGT.....AGCAACACCGCGCGCCCTTC 258

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb\_bal:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_ph:\*  
6: gb\_pl1:\*  
7: gb\_pl2:\*  
8: gb\_pri:\*  
9: gb\_pr2:\*  
10: gb\_pr3:\*  
11: gb\_ro:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: em\_fun:\*  
15: em\_hum1:\*  
16: em\_hum2:\*  
17: em\_in:\*  
18: em\_om:\*  
19: em\_or:\*  
20: em\_ov:\*  
21: em\_pat:\*  
22: em\_ph:\*  
23: em\_pl:\*  
24: em\_ro:\*  
25: em\_sts:\*  
26: em\_sy:\*  
27: em\_un:\*  
28: em\_v1:\*  
29: gb\_ba3:\*  
30: gb\_in1:\*  
31: gb\_in2:\*  
32: gb\_in3:\*  
33: gb\_pl3:\*  
34: gb\_pr4:\*  
35: em\_bal:\*  
36: em\_ba2:\*  
37: em\_htg1:\*  
38: em\_htg2:\*  
39: em\_htg3:\*  
40: em\_htg4:\*  
41: em\_htg5:\*  
42: em\_htg6:\*  
43: em\_htg7:\*

44: em\_htg8:\*  
45: em\_htg9:\*  
46: em\_htg10:\*  
47: em\_hum3:\*  
48: em\_hum4:\*  
49: em\_hum5:\*  
50: em\_hum6:\*  
51: gb\_pr5:\*  
52: gb\_pr6:\*  
53: gb\_pr7:\*  
54: gb\_htg1:\*  
55: gb\_htg2:\*  
56: gb\_htg3:\*  
57: gb\_htg4:\*  
58: gb\_htg5:\*  
59: gb\_htg6:\*  
60: gb\_htg7:\*  
61: gb\_htg8:\*  
62: gb\_htg9:\*  
63: gb\_htg10:\*  
64: gb\_htg11:\*  
65: gb\_htg12:\*  
66: gb\_htg13:\*  
67: gb\_htg14:\*  
68: gb\_htg15:\*  
69: gb\_htg16:\*  
70: gb\_htg17:\*  
71: gb\_htg18:\*  
72: gb\_htg19:\*  
73: gb\_htg20:\*  
74: gb\_htg21:\*  
75: gb\_htg22:\*  
76: gb\_htg23:\*  
77: gb\_sts1:\*  
78: gb\_sts2:\*  
79: gb\_vil:\*  
80: gb\_v12:\*  
81: gb\_pat1:\*  
82: gb\_pat2:\*  
83: em\_htg0:\*  
84: gb\_htg24:\*  
85: gb\_pr8:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	258	100.0	477	10	AF083421	AF083421 Homo sapi
2	258	100.0	1230	53	HSD1F2	Y14551 Homo sapien
3	258	100.0	1309	10	AF039067	AF039067 Homo sapi
4	258	100.0	1693	10	AF071596	AF071596 Homo sapi
5	258	100.0	1864	53	HSPRG1	X96438 H.sapiens P
6	258	100.0	44118	9	AC006165	AC006165 Homo sapi
7	258	100.0	90244	8	AB023051	AB023051 Homo sapi
8	258	100.0	200000	51	AF000512	AF000512 Homo sapi
9	256.4	99.4	1223	85	S81914	S81914 TEX-1-radia
10	154.4	59.8	1758	11	RNPRG1	X96437 R.norvegicu
11	151.2	58.6	203476	70	AC074150	AC074150 Mus muscu
12	149.6	58.0	1938	11	MMGLY96	X67644 M.musculus
13	139.2	54.0	187868	62	AC022301	AC022301 Mus muscu
14	44	17.1	297	81	A75448	A75448 Sequence 11
15	44	17.1	297	81	A78427	A78427 Sequence 11
16	43.2	16.7	115468	57	AC011820	AC011820 Homo sapi
17	43	16.7	7218	81	I66494	I66494 Sequence 14
18	40.2	15.6	34725	29	SC1B2	AL356812 Streptomy
19	39.8	15.4	3897	2	MXU40656	U40656 Myxococcus
20	39.6	15.3	33818	2	MYCY159	Z83863 Mycobacteri
21	39.4	15.3	2003	85	HUMDAFC1	M64356 Human decay

BASE COUNT 66 a 117 c 61 g 52 t 1 others  
ORIGIN

Query Match 12.3%; Score 151; DB 81; Length 297;  
Best Local Similarity 76.0%; Pred No. 1.6e-24;  
Matches 212; Conservative 0; Mismatches 65; Indels 2; Gaps 2;

Qy 1 ATGTGTCACCTCGCAGCTGCCACCCGACCATGACCATCTGTCAGGCCCCGACCCCGGCC 60  
Db 16 ATGGAACACTCTCGCGCTGCCACTCCCCCATGACCACTTACAGGCCCTAACCCAGCC 75

Qy 61 CCCTCCACCATCCCGGACCCCGGGGGCTCCGGTCTGAGATCTTCACCTTCGACCCCT 120  
Db 76 CCTCAGGTATCTGCCAGCCACAATA-GCTGCCGCGCTGAGATCTCCACCTCAGACCCCT 134

Qy 121 CTCCCGAGCCCGCAGGCGCCCTGCCGGGCGCCCGCCGCTCTCGGGGCACCCGAAAG 180  
Db 135 CTCCGGAAGCCACAGGGGCGCCCTGCCGGGACCCCGAGCTTC-CTCGAGGCACCAAG 193

Qy 181 CGCAGCCGACGGTTCTTACCTCGAGTGTCCGGGCGCCAGTGTCCAGTCTGAGGAACCG 240  
Db 194 CGCAGTTCCAGGAATCTCTACCTAGAGTATCCGGGCGCCAGTTGCCAGTTAAGGACCA 253

Qy 241 AACCCAGCCAAAGGCTTCTCTTCTGCTGCTACCAATC 279  
Db 254 AACCTGCCAGAGGCTTCTTTTCGGGTTGATCGCCATC 292

Search completed: January 30, 2001, 19:23:39  
Job time: 25905 sec



```

Oy 312 TGTGCGCGGCCCTGCTCCAGAGACGCCCTTAAGCGCGCATCC-----T 359
Db 1132 TGTGTCGACGCCCTGCTCCGAGGATGTACACGCCGCTGACACCTGAGCCCATTC 1191
Oy 360 GGGCGCCACCCCTGTGTGTCGCCCTGCTCGAGCCCTTAACTGACTTCGGAGCCCTCGGA 419
Db 1192 TGGGCCCATTAAGTGGCGGCCCTGCTCGAGCCTTGAACCTGACCTCGGAGTCTCGGA 1251
Oy 420 CTAGGCTCTGGACTGACACTTCTCTCCAGCAACACCCCGCGCTTCTAACTGTGACT 479
Db 1252 CTATGCGCTGGATCTTAAAGCTTCTCTAGCAACATCCGCGCGCTTCTTAAAGCGAT 1311
Oy 480 CCGCGCACTCCCAAAAGAAATCCGAAACACCAAGAACACCCAGCGCTACTGTGTC 539
Db 1312 GCTC-----ACAAGTCGGAAGAACAAAGCCACCATGGATGGTACTGTGTC 1358
Oy 540 GCGAGACGATATCCCAACTGGGACTTCCGAGGCAACTTGAATCTAGAACACTACAGCG 599
Db 1359 GAGAGACGATATCCCAACTGGGATTTCTTAAGCAAGCTTAATCTAGAACACTACCGCCA 1418
Oy 600 AGAGCCACCCGGTGTG 618
Db 1419 AGAGACACCGCGGTCTG 1437

RESULT 14
AC022301
LOCUS
DEFINITION Mus musculus clone RP23-329K24, *** SEQUENCING IN PROGRESS ***, 59
unordered pieces.
AC022301
VERSION AC022301.9 GI:8493517
KEYWORDS HTG; HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 187868)
Dedrich,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Metzker,M.L., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Gorell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogue,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstein,G.,
Worley,K. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 187868)
Worley,K.C.
Direct Submission
Submitted (29-JAN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 13, 2000 this sequence version replaced gi:8248590.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MAAH
Center clone name: RP23-329K24
----- Summary Statistics
Sequencing vector: M13; L08821

```

```

Chemistry: Dye-primer Bodipy: 82% of reads
Chemistry: Dye-terminator Big Dye: 18% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 133377 bases at least Q40
Consensus quality: 159061 bases at least Q30
Consensus quality: 169688 bases at least Q20
Estimated insert size: 166938; sum-of-contigs estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 7154: contig of 7154 bp in length
* 7155 7254: gap of unknown length
* 7255 14340: contig of 7086 bp in length
* 14341 14440: gap of unknown length
* 14441 22916: contig of 8476 bp in length
* 22917 23016: gap of unknown length
* 23017 30834: contig of 7818 bp in length
* 30835 30934: gap of unknown length
* 30935 37933: contig of 7059 bp in length
* 37934 38093: gap of unknown length
* 38094 45511: contig of 7418 bp in length
* 45512 51842: contig of 6231 bp in length
* 51843 51942: gap of unknown length
* 51943 58572: contig of 6630 bp in length
* 58573 58672: gap of unknown length
* 58673 62718: contig of 4046 bp in length
* 62719 62818: gap of unknown length
* 62819 67974: contig of 5156 bp in length
* 67975 68074: gap of unknown length
* 68075 72883: contig of 4809 bp in length
* 72884 72983: gap of unknown length
* 72984 76726: contig of 3743 bp in length
* 76727 76826: gap of unknown length
* 76827 79631: contig of 2805 bp in length
* 79632 79731: gap of unknown length
* 79732 82858: contig of 3127 bp in length
* 82859 82958: gap of unknown length
* 82959 86054: contig of 3096 bp in length
* 86055 86154: gap of unknown length
* 86155 90283: contig of 4129 bp in length
* 90284 90383: gap of unknown length
* 90384 94598: contig of 4215 bp in length
* 94599 94698: gap of unknown length
* 94699 99014: contig of 4316 bp in length
* 99015 103295: contig of 4181 bp in length
* 103296 103395: gap of unknown length
* 103396 106698: contig of 3303 bp in length
* 106699 106799: gap of unknown length
* 106799 109658: contig of 2859 bp in length
* 109658 109757: gap of unknown length
* 109758 113392: contig of 3635 bp in length
* 113393 113492: gap of unknown length
* 113493 117011: contig of 3519 bp in length
* 117012 117111: gap of unknown length
* 117112 120036: contig of 2925 bp in length
* 120037 120136: gap of unknown length
* 120137 122449: contig of 2313 bp in length
* 122450 122549: gap of unknown length
* 122550 124276: contig of 1726 bp in length
* 124276 124375: gap of unknown length
* 124376 126296: contig of 1921 bp in length
* 126297 126396: gap of unknown length
* 126397 129180: contig of 2784 bp in length
* 129181 129280: gap of unknown length

```

```
* 95025 101188: contig of 6164 bp in length
* 101189 101288: gap of unknown length
* 101289 108925: contig of 7637 bp in length
* 108926 109025: gap of unknown length
* 109026 119738: contig of 10713 bp in length
* 119739 119838: gap of unknown length
* 119839 129876: contig of 10038 bp in length
* 129877 129976: gap of unknown length
* 129977 143243: contig of 13267 bp in length
* 143244 143343: gap of unknown length
* 143344 161164: contig of 17821 bp in length
* 161165 161264: gap of unknown length
* 161265 180316: contig of 19052 bp in length
* 180317 180416: gap of unknown length
* 180417 203476: contig of 23060 bp in length.
FEATURES
    Location/Qualifiers
    source
    1..203476
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="RP23-128D3"
    /clone_lib="RPCI mouse BAC library 23"
BASE COUNT 51748 a 48114 c 48138 g 51580 t 3296 others
ORIGIN
Query Match 17.1%; Score 210.6; DB 70; Length 203476;
Best Local Similarity 62.4%; Pred. No. 2.le-38;
Matches 461; Conservative 0; Mismatches 144; Indels 134; Gaps 3;
Oy 1 ATGTCTACTCTCGAGCTGCCACCGGAGCTGACCATGACCATCTGCGAGGCCCCGAGCCGCC 60
Db 92556 ATGTGCCACTCGCGAACCATCTCCACCATGACTGCGCTGAGGGGCCCTTCTCCAGCT 92615
Oy 61 CCTCCACCATCTCCGAGCCCGCGGGCTCCGCTCTGAGATCTTCCACCTTCGACCT 120
Db 92616 CCTCCACCGGCGGAACTCGGGGGGCTCTGTCTCCGAAATTTTACCTTCGACCT 92675
Oy 121 CTCCTGGAGCCGAGCGGCGCTTCGCGGGGCGCCCGAGCGCTTCGCGGCGACCGAAAG 180
Db 92676 CTCCTGGAGCGGCGGCTTCGCGGCGGCTTCGCGGCGGCTTCGCGGCGGCGGCGG 92735
Oy 181 CGACCGCGAGGGTTCTTACCTCGAGT----- 209
Db 92736 CGACCGCGAGGGTTCTTACCTCGAGTGGTGTAGTATCGCGCGAGTGGGATCAGGAGG 92795
Oy 210 ----- 209
Db 92796 TCGCGTGCCTTGAACCTGTAGTTAACAACTAGGCGGAGGACCTTCGATCTGACGTTT 92855
Oy 210 -----GTCGCGGCGGCGGCTTCGAGTTCGAGAACCGACCGACCA 251
Db 92856 CCTCTTTTATCTGCTCAGGTCGCGGCGGCTTACCAACCGAGGACCAACATTGCCAA 92915
Oy 252 AAGGCTCTCTTCTGCTGCTCAGCATCTCTTCTGCGAGATCTGTGCTGAGAGGG 311
Db 92916 GAGGCTCTCTTCTGCTGCTCAGCATCTCTTCTGCGAGATCTGTGCTGAGAGGG 92975
Oy 312 TGTGCGGCGGCGGCTTCTGCTGCGAGGAGCGGCGCTTAAACCGCATCCC-----T 359
Db 92976 TGTGTCGAGCGGCGGCTTCTGCTGCGAGGAGTGTCTACCGCGCGCTTGACCTGAGCCATTTC 93035
Oy 360 GCGCGCGACCTGTGTCGCGGCTTCTGCGAGGCTTAACTGTGCTGCGAGCGGCTCGGA 419
Db 93036 TGTGTCGAGCGGCGGCTTCTGCTGCGAGGAGTGTCTACCGCGCGCTTGACCTGAGCCATTTC 93095
Oy 420 CTACGCTCTGAGCTCAGCACTTCTCTCCAGCAACACCGCGGCGCTTCTAACTGTGACT 479
Db 93096 CTATCGGCTGATCTTAAAGCTTCTTCCAGCAACATCGCGGCGCTTCTAAACGCGATG 93155
Oy 480 CCCCCTACTCCCAAAAGATCCGAAACCAACCAAGAACACGAGCGGTACCTGTGTGC 539
Db 93156 GGTG-----ACATCCGAGAACAAAGGACCATGGATGGGTACCTGTGTGC 93202
```

```
Oy 540 GCGAGAGCGTATCCCAACTCGGACTTCCGAGGCAACTTGAATCAGAACACTACACGG 599
Db 93203 GAGAGACGATATCCCAACTCGGACTTCTAAGGCAACCTTAACTCAGAACACTACCCCA 93262
Oy 600 AGACGCCACCGCGGCTTGTG 618
Db 93263 AGAGACCGCGGCTGCTG 93281
RESULT 13
MMGLY96 1938 bp mRNA ROD 11-MAR-1993
LOCUS M.musculus gly96 mRNA.
DEFINITION X67644
ACCESSION X67644
VERSION X67644.1 GI:287803
KEYWORDS gly96 gene; glycosylated protein.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1938)
AUTHORS Lau, L.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1992) L. Lau, University of Illinois College of
Med., Dept. of Genetics, M/C 669, 808 South Wood Street, Chicago,
IL 60612, USA
REFERENCE 2 (bases 1 to 1938)
AUTHORS Charles, C.H., Yoon, J.K., Simske, J.S. and Lau, L.F.
TITLE Genomic structure, cDNA sequence, and expression of gly96, a growth
factor-inducible immediate-early gene encoding a short-lived
glycosylated protein
JOURNAL Oncogene 8 (3), 797-801 (1993)
MEDLINE 93173526
FEATURES Location/Qualifiers
    source
    1..1938
    /organism="Mus musculus"
    /strain="BALB/c"
    /db_xref="taxon:10090"
BASE COUNT 390 a 525 c 503 g 520 t
ORIGIN
Query Match 17.0%; Score 209; DB 11; Length 1938;
Best Local Similarity 62.2%; Pred. No. 6.4e-38;
Matches 460; Conservative 0; Mismatches 145; Indels 134; Gaps 3;
Oy 1 ATGTGTCACTCTCCGAGCTGCCACCGGAGCTTCCGAGTCTTCTGAGATCTTCCACCTTCGACCT 60
Db 712 ATGTGCCACTCGCGAACCATCTCCACCATGACTGCGCTGAGGGGCCCTTCTCCAGCT 771
Oy 61 CCTTCCACCATCTCCGAGGAGCGGCGGCTTCCGAGTCTTCTGAGATCTTCCACCTTCGACCT 120
Db 772 CCTTCCACCGGCGGAGAACTCCGCGGGGCTCTGTCTCCGAAATTTTACCTTCGACCT 831
Oy 121 CTCCTGGAGCCGCGGCGGCGGCTTCCGCGGCGGCGGCGGCTCTCGCGGCGGCGGCGGAG 180
Db 832 CTCCTGGAGCGGCGGCTTCTGCTCCACCGCGGCTTGAACACTTCTCGCGGCGGCGGAGAA 891
Oy 181 CGCAGCGCGAGGTTCTTACCTCGAGT----- 209
Db 892 CGCAGCGCGAGGTTCTTACCTCGAGTGGTGTAGTATCGCGCGGAGTGGGATCAGGAGG 951
Oy 210 ----- 209
Db 952 TCGCGTGCCTTGAACCTTGTAGTTAACAACTAGGAGCGGAGCTTTCGATCTTCACGTTT 1011
Oy 210 -----GTCGCGGCGGCGGCTTCCGAGTTCGAGAACCGAACCGACCA 251
Db 1012 CCTCTTTTATCTGCTCAGGTCGCGGCGGAGTACCAACCGAGGACCAACATTGCCAA 1071
Oy 252 AAGCTTCTCTTCTGCTGCTCAGCATCTTCTGCGAGGCTCTGATGCTGATGCTGAGAGGG 311
Db 1072 GAGGCTCTCTTCTGCTGCTCAGCATCTTCTGCGAGATTTTGTGCTGATGCTGAGAGGG 1131
```



```
Db 67 CCTCCACCATCCGGGAGCCCGGGGGCTCCGGTCCCTGAGATCTTCACCTTCGACCCCT 126
Qy 121 CTCCTGGAGCCCGCAGCGCCCTCGCGGGCGCCCGCAGCGCTCTCGGGGACCGAAG 180
Db 127 CTCCTGGAGCCCGCAGCGCCCTCGCGGGCGCCCGCAGCGCTCTCGGGGACCGAAG 186
Qy 181 CCGAGCCGAGGCTTCTCTACCTCGAGTGGTCCGGCGCCAGCTGCCAGTCGAGAACCG 240
Db 187 CCGAGCCGAGGCTTCTCTACCTCGAGTGGTCCGGCGCCAGCTGCCAGTCGAGAACCG 246
Qy 241 AACCCAGCCAAAGGCTTCTCTTGTGTCACCATGCTTCTGCGAGATCTTCGATG 300
Db 247 AACCCAGCCAAAGGCTTCTCTTGTGTCACCATGCTTCTGCGAGATCTTCGATG 306
Qy 301 GCTGAGAGGGTGTCCGGCGCCCTGCTCCAGAGGAGCCCTTAACGCCGATCCCTG 360
Db 307 GCTGAGAGGGTGTCCGGCGCCCTGCTCCAGAGGAGCCCTTAACGCCGATCCCTG 366
Qy 361 GCGCCACCCCTGTGTCGCCCGCTCGAGCCCTTTAACTGACCTCGGAGCCCTCGGAC 420
Db 367 GCGCCACCCCTGTGTCGCCCGCTCGAGCCCTTTAACTGACCTCGGAGCCCTCGGAC 426
Qy 421 TAGCCTCTGAGCTCAGCACTTTCTCTCCAGCAACCCCGCGCTTCTAA 471
Db 427 TAGCCTCTGAGCTCAGCACTTTCTCTCCAGCAACCCCGCGCTTCTAA 477

RESULT 10
HSA227914 343 bp mRNA PRI 17-DEC-1998
LOCUS Homo sapiens partial mRNA; ID YG40-2.
DEFINITION AJ227914
ACCESSION AJ227914
VERSION AJ227914.1 GI:3183967
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 343)
AUTHORS Dominguez, O.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-1998) Dominguez O., Immunologia, Hospital Germans
Trias i Pujol, 08916 Badalona, and Research Center, Almirall
Prodesfarms, 08024 Barcelona, Spain
REFERENCE 2 (bases 1 to 343)
AUTHORS Dominguez, O., Ashhab, Y., Sabater, L., Belloso, E., Caro, P. and
Pujol-Borrell, R.
TITLE Cloning of ARE-containing genes by AU-motif-directed display
JOURNAL Genomics 54 (2), 278-286 (1998)
COMMENT 99047534
FEATURES
source
1..343
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="adult"
/cell_type="T lymphocyte"
/cell_line="T12.12"
/clone="PH19-20, PH19-21"
<1..>343
/gene="DIF-2"
1..343
/gene="DIF-2"

mRNA 72 a 67 c 114 g 85 t 5 others
gene
BASE COUNT
ORIGIN

Query Match 24.3%; Score 298.8; DB 52; Length 343;
Best Local Similarity 97.1%; Pred. No. 1.4e-58;
Matches 333; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

Qy 862 GCTTTGCGGACTGGTGAGAGCAGGACCTCGAACTCGGCGAAAGTAGGAGAGAAATG 921
|||||
```

```
Db 1 GCTTTGCGGACTGGTGAGAGCAGGAC-CCITGAACCTCCGCAAGTAGGACAAG-AAATG 58
Qy 922 GGGAGGACTCGGCTGGGGAGGAGCTCCGGCTGGGATGAAGTCTGTGGTGGGTCTGTAA 981
Db 59 GGGAGGACTCGGCTGGGGAGGAGCTCCGGCTGGGATGAAGTCTGTGGTGGGTCTGTAA 118
Qy 982 GTTTAGGAGGTGACTTGCATCTCCAGCATCTCAACTCCGCTGTCTACTGTGTGAGACTT 1041
Db 1119 G-TTAGGAGGTGACTTGCATCTCCAGCATCTCAACTCCGCTGTCTACTGTGTGAGACTT 177
Qy 1042 CGGGGAGCCATTAGGAATGAGATCCGTGAGATCTTCCATCTTCTTGAAGTCGGCTTTAG 1101
Db 178 CGGGGAGCCATTAGGAATGAGATCCGTGAGATCTTCCATCTTCTTGAAGTCGGCTTTAN 237
Qy 1102 GSTGCTCGGAGGTAGAGGTTGGGGTTGGTGGCTGTCAAGGAGGACTGTGAGATC 1161
Db 238 GGTGCTCGGAGGTAGAGGTTGGGGTTGGTGGCTGTCAAGGAGGACTGTGAGATC 297
Qy 1162 GCCTAGTATGTTCTGTGAACACAAATAAATTGATTTACTGTC 1204
Db 298 GCCTANTATGTTCTGTGAACACAAATAAATTGATTTACTGTC 340

RESULT 11
RNPRG1 1758 bp DNA ROD 03-SEP-1996
LOCUS R.norvegicus PRG1 gene.
DEFINITION X96437
ACCESSION X96437.1 GI:1515318
VERSION X96437.1
KEYWORDS PRG1 gene.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 1758)
AUTHORS Schaffer, H., Trauzold, A., Siegel, E.G., Polsch, U.R. and Schmidt, W.E.
TITLE PRG1: a novel early-response gene transcriptionally induced by
pituitary adenylate cyclase activating polypeptide in a pancreatic
carcinoma cell line
JOURNAL Cancer Res. 56 (11), 2641-2648 (1996)
MEDLINE 96221139
REFERENCE 2 (bases 1 to 1758)
AUTHORS Trauzold, A.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1996) A. Trauzold, Laboratory of Molecular
Gastroenterology, 1st Dept. of Medicine, University of Kiel,
Schittenhelmstrasse 12, Kiel, D-24105, FRG
FEATURES
source
1..1758
Location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/cell_line="AR4-2J"
590..1279
/gene="PRG1"
590..1279
/gene="PRG1"

BASE COUNT 358 a 475 c 470 g 455 t
ORIGIN

Query Match 19.7%; Score 242.4; DB 11; Length 1758;
Best Local Similarity 62.1%; Pred. No. 1.3e-45;
Matches 589; Conservative 0; Mismatches 181; Indels 178; Gaps 6;

Qy 1 ATGTGTCACTCTCGAGCTGCCACCGCATGACCATCTTCAGCCCGCCCGCCCGCCGCC 60
|||||
Db 590 ATGTGCGATTCCGTTAACCACCTCCACACCATGACTGCGCTGAGGGCTCCGGCGCCAGTT 649
Qy 61 CCTCCACCATCCCGGAGCCCGGGGCTCCGGTCTCTGAGATCTTCACTTCGACCT 120
|||||
Db 650 CCTCCACCGCCCGGAACTCCGGGCTCCGGTCCCGGAAATTTTCACTTCGACCCC 709
Qy 121 CTCCCGGAGCCCGCAGCGCCCTCGCGGGCGCCCGCTTCGCGGGCGCCCGCCGAAAG 180
```







Qy	250	AAAAGCGCTCTCTTTCTGCTGCTCACCATCCTCTTCTCCAGATCCTGATGGCTGAAGAG	309
Db	361	AAAAGGCTTCTCTTTCTGCTGCTCACCATCCTCTTCTGCGCAGATCCTGATGGCTGAAGAG	420
Qy	310	GGTGTGCGGCGCCCTGCTCCTCCAGAGAGCGCCCTAACCGCGCATCCCTGGGCGCCACC	369
Db	421	GGTGTGCGGCGCCCTGCTCCTCCAGAGAGCGCCCTAACCGCGCATCCCTGGGCGCCACC	480
Qy	370	CCTGTGCTCCCGCTCTCGAGCGCTTTAATCTGACTTCTCGAGCGCTCGACTAGGCTCTG	429
Db	481	CCTGTGCTCCCGCTCTCGAGCGCTTTAATCTGACTTCTCGAGCGCTCGACTAGGCTCTG	540
Qy	430	GACCTCAGCACTTTCTCTCCAGCAACACCCCGCGCTTCTAATCTGACTTCTCCCGCACTC	489
Db	541	GACCTCAGCACTTTCTCTCCAGCAACACCCCGCGCTTCTAATCTGACTTCTCCCGCACTC	600
Qy	490	CCAAAANGAATCGGAAACACCAAGAAACACCAAGCGCTGCTGTCGCGAGAGCGT	549
Db	601	CCAAAANGAATCGGAAACACCAAGAAACACCAAGCGCTGCTGTCGCGAGAGCGT	660
Qy	550	ATCCCCAATGGGACTTCCGAGGCACTTCAACTCAGAACTACAGCGGAGAGCGCCACC	609
Db	661	ATCCCCAATGGGACTTCCGAGGCACTTCAACTCAGAACTACAGCGGAGAGCGCCACC	720
Qy	610	CGGTGCTTGAGCGCGGACCGAGCGCACAGAGACCGGCGCATAGAGACCGAGGCAACAG	669
Db	721	CGGTGCTTGAGCGCGGACCGAGCGCACAGAGACCGGCGCATAGAGACCGAGGCAACAG	780
Qy	670	CCAGCTGGGGCTAGGCGCGGCTGGGAAGGAGAGCGTCTTAATTTATTTCTTATGCTCC	729
Db	781	CCAGCTGGG--CTAGGCGCGTGGGAAGGAGAGCGTCTTAATTTATTTCTTATGCTCC	838
Qy	730	TAATTAATATTTATATCTATTTATGTAAGTCTCTCTAGTGTGATGGAGATGTGTACGTAAT	789
Db	839	TAATTAATATTTATATCTATTTATGTAAGTCTCTCTAGTGTGATGGAGATGTGTACGTAAT	898
Qy	790	ATTATTTTAACTTATCAAGGGTGTGAGATGTTCCCTCTGCTGTAATTCAGGCTCTCTT	849
Db	899	ATTATTTTAACTTATCAAGGGTGTGAGATGTTCCCTCTGCTGTAATTCAGGCTCTCTT	958
Qy	850	GGTATTTATTTGAGCTTTGTGGGACTGTGTGGAAGCAGACACCTTGGAACTCGGCAAGTA	909
Db	959	GGTATTTATTTGAGCTTTGTGGGACTGTGTGGAAGCAGACACCTTGGAACTCGGCAAGTA	1018
Qy	910	GGAGAGAAATGGGAGGACTCGGCTGGGGAGGAGCGTCCCGGCTGGGATGAAGTCTGGT	969
Db	1019	GGAGAGAAATGGGAGGACTCGGCTGGGGAGGAGCGTCCCGGCTGGGATGAAGTCTGGT	1078
Qy	970	GGTGGGTCGTAACTTTAGGAGGTGACTTGCATCCTCCAGCATCTCAACTCCGCTGCTGTAC	1029
Db	1079	GGTGGGTCGTAACTTTAGGAGGTGACTTGCATCCTCCAGCATCTCAACTCCGCTGCTGTAC	1138
Qy	1030	TGTGTGAGACTTCGGGGAGCAATTAGGAATGAGATCCGTGAGATCTTCCATCTTTGA	1089
Db	1139	TGTGTGAGACTTCGGGGAGCAATTAGGAATGAGATCCGTGAGATCTTCCATCTTTGA	1198
Qy	1090	AGTCGGCTTTAGGGTGGCTCGAGGTAGAGGGTTGGGGTTGGTGGGCTGTACGAGAGCG	1149
Db	1199	AGTCGGCTTTAGGGTGGCTCGAGGTAGAGGGTTGGGGTTGGTGGGCTGTACGAGAGCG	1258
Qy	1150	ACTGTCGAGATCCCTAGTATGTTCTGTGAACACAAATAAATTTGATTTAC	1200
Db	1259	ACTGTCGAGATCCCTAGTATGTTCTGTGAACACAAATAAATTTGATTTAC	1309
RESULT	7		
AF071596			
LOCUS	1693 bp	DNA	10-NOV-1998
DEFINITION	Homo sapiens apoptosis inhibitor (IEX-1L) gene, complete cds.		
ACCESSION	AF071596		
VERSION	AF071596.1	GI:3851531	
KEYWORDS	human.		
SOURCE			

ORGANISM	Homo sapiens		
REFERENCE	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Wu, M.X., Ao, Z., Prasad, K.V., Wu, R. and Schlossman, S.F.		
TITLE	IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival		
JOURNAL	Science 281 (5379), 998-1001 (1998)		
MEDLINE	98369175		
REFERENCE	2 (bases 1 to 1693)		
AUTHORS	Wu, M.X. and Ao, Z.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-JUN-1998) Tumor Immunology, Dana Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA		
FEATURES	Location/Qualifiers		
source	1..1693		
organism	"Homo sapiens"		
db_xref	taxon:9606		
359..1693			
gene	/gene="IEX-1L"		
product	"apoptosis inhibitor"		
359..1693			
gene	/gene="IEX-1L"		
385..966			
note	"IEX-1L"		
note	"immediate early response protein"		
codon_start=1			
product	"apoptosis inhibitor"		
protein_id="AAC72344.1"			
db_xref="GI:3851532"			
translation="MCHSRSCPTMTLQAPTPAPSTIPGRKSGPVEITFDPLDEP			
AAAPGRPSASGRKRRSRVSLPSSGTYRRSGHSCALPWSHGTRLOSLLDITCII.			
FCPLVRLPVEENPARKLLLTIVFCILMAEGVPAPLPPEADAPNAASLAPT			
PVSPVLEPENLTSEPSYALDLSFLQHPAAF"			
BASE COUNT	348 a	529 c	456 g
ORIGIN	360 t		
Query Match	85.4%	Score 1048.6;	DB 10; Length 1693;
Best Local Similarity	91.1%	Pred. No. 2.6e-231;	
Matches 1194;	Conservative	0; Mismatches	4; Indels 113; Gaps 3;
Qy	1	ATGTGTCACCTCTCGAGCTGCCACCGGACCATCACCATCTCGAGGCCCCCGCCGCC	60
Db	385	ATGTGTCACCTCTCGAGCTGCCACCGGACCATCACCATCTCGAGGCCCCCGCCGCC	444
Qy	61	CCCTCCACCATCCCGGGACCCCGGGGCTCCGCTCCTGAGATCTTACCTTCGACCTT	120
Db	445	CCCTCCACCATCCCGGGACCCCGGGGCTCCGCTCCTGAGATCTTACCTTCGACCTT	504
Qy	121	CTCCGAGAGCCGACGCGGCTCTGCGGGGCCCCGAGGCGCTCTCGGGGACCCGAAAG	180
Db	505	CTCCGAGAGCCGACGCGGCTCTGCGGGGCCCCGAGGCGCTCTCGGGGACCCGAAAG	564
Qy	181	CGAGCGGAGGGT-TCTCTACCTCGAGTG-----	210
Db	565	CGAGCGGAGGGTCTCTTACCTCGAGTGAGTATGCGCGAAAGTGGGATTCGCGG	624
Qy	211	-----	210
Db	625	TGTCGCTGCCCTGGAGTCACATGGGGAACGACCCGACTCCAGAGCCCTCGACCTGACCTGT	684
Qy	211	-----GTCCGGGCGGAGCTGCCAGTTCGAGGAGAACCGAACCCAGCC	249
Db	685	CTCCTGTTTGTCTCCCTTAGTCTCGGCGGAGCTGCCAGTCGAGGAACGAAACCCAGCC	744
Qy	250	AAAAGCTTCTCTTTCTGCTCACCATCGTCTTCTGCCAGATCTCTGATGGCTGAAGAG	309
Db	745	AAAAGCTTCTCTTTCTGCTCACCATCGTCTTCTGCCAGATCTCTGATGGCTGAAGAG	804
Qy	310	GGTGTGCGGCGGCTGCTCTCCAGAGAGCGCCCTTAACGCGCATCTCTGCGGCGCCACC	369
Db	805	GGTGTGCGGCGGCTGCTCTCCAGAGAGCGCCCTTAACGCGCATCTCTGCGGCGCCACC	864



```

Db 61518 CTAATTAATATTATATGTAATTTATGTAAGTCTCTCTAGGTGATGGAGATGTTAGCTAA 61577
Qy 789 TATTTTATTTTAACTTATCCAGGGTGTGAGATGTTCCCTCTGCTGCTAAATCAGGTCTCT 848
Db 61578 TATTTTATTTTAACTTATCCAGGGTGTGAGATGTTCCCTCTGCTGCTAAATCAGGTCTCT 61637
Qy 849 TGGTATTTATTTAGCTTTGTGGGACTGTGGAAGCAGGACACCTGGAACTCGCGCAAGT 908
Db 61638 TCGTATTTTATTTAGCTTTGTGGGACTGTGGAAGCAGGACACCTGGAACTCGCGCAAGT 61697
Qy 909 AGGAGAGAAATGGGGAGACTTCGGGTGGGGAGAGACTTCGGGTGGGGATGAAGTCTGG 968
Db 61698 AGGAGAGAAATGGGGAGACTTCGGGTGGGGAGAGACTTCGGGTGGGGATGAAGTCTGG 61757
Qy 969 TGGTGGCTCGTAAGTTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCGCTGTGCTA 1028
Db 61758 TGGTGGGTGGTAAGTTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCGCTGTGCTA 61817
Qy 1029 CTGTGTAGACTTCGGCGGACCATTAGGAATGAGATCGGTGAGATCCTTCCATCTTCTTG 1088
Db 61818 CTGTGTAGACTTCGGCGGACCATTAGGAATGAGATCGGTGAGATCCTTCCATCTTCTTG 61877
Qy 1089 AAGTCGCCCTTTAGGTGCTCCGAGGTAGACGGTTGGGGTTGGTGGCTGTCTACGGAGC 1148
Db 61878 AAGTCGCCCTTTAGGTGCTCCGAGGTAGACGGTTGGGGTTGGTGGCTGTCTACGGAGC 61937
Qy 1149 GACTGTCCAGATCCCTAGTATGTTCTGTGAACAATAAATGATTTACTGTC 1204
Db 61938 GACTGTCCAGATCCCTAGTATGTTCTGTGAACAATAAATGATTTACTGTC 61993

RESULT 5
LOCUS AP000512 200000 bp DNA PRI 30-MAR-2000
DEFINITION Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region,
section 11/20.
ACCESSION AP000512
VERSION AP000512.1 GI:5926699
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Shina,S., Tamiya,G., Oka,A. and Inoko,H.
Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region
PUBLISHED Only in DataBase (1999) in press
2 (bases 1 to 200000)
Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
Direct Submission
Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases, Miki
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail:mika@tokyo.jst.go.jp,
URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)
This sequence is conducted by Tokai University as a JST sequencing
Team.
COMMENT
Principal Investigator: Hidetoshi Inoko Ph.D
Phone:+81-463-93-1121, Fax:+81-463-94-8884,
The sequence is submitted by Human Genome Sequencing in ALIS
project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences, please visit our
sequence archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/top.
html) or send email to webmaster@www-alis.tokyo.jst.go.jp.
FEATURES
Location/Qualifiers
1. 200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"

```

```

STS 101440..101648
/note="s1564264:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4566033"
complement(114197..114328)
/standard_name="D6S2115"
/note="SHGC-13451:The location is between each flanking
site of PCR primers."
/db_xref="GDB:738998"
complement(114300..114430)
/note="WI-18721:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4574312"
117236..117362
/note="SGC34396:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4580073"
complement(117712..117836)
/standard_name="D6S2041"
/note="SHGC-10364:The location is between each flanking
site of PCR primers."
/db_xref="GDB:735115"
complement(117961..118093)
/note="SGC32809:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4584619"
142184..142339
/note="K1AA0170:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4586652"
complement(142215..142343)
/note="WI-15445:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4581086"
complement(153977..154114)
/note="SGC34551:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4582352"
complement(165191..165295)
/standard_name="D6S1376"
/note="downstream/upstream:The location is between each
flanking site of PCR primers."
/db_xref="GDB:580497"
188925..188989
/note="pml216:The location is between each flanking site
of PCR primers."
/db_xref="GDB:594325"
complement(197650..197804)
/note="Cda01e04:The location is between each flanking site
of PCR primers."
/db_xref="GDB:443319"
BASE COUNT 52458 a 48336 c 49153 g 50053 t
ORIGIN

Query Match 88.0%; Score 1080.4; DB 51;
Best Local Similarity 91.4%; Pred. No. 9.6e-239;
Matches 1203; Conservative 0; Mismatches 1; Indels 112; Gaps 1;

Qy 1 ATGTGTCACTCTCGAGCTGCCACCGACCATGACCATCTCGAGGCTCGGGTCTTGAGATCTTTCAGCTTCGAGCCCT 120
Db 97663 ATGTGTCACTCTCGAGCTGCCACCGACCATGACCATCTCGAGGCTCGGGTCTTGAGATCTTTCAGCTTCGAGCCCT 97722
Qy 61 CCTCCACCATCCCGGACCCCGCGGGCTCGGGTCTTGAGATCTTTCAGCTTCGAGCCCT 120
Db 97723 CCTCCACCATCCCGGACCCCGCGGGCTCGGGTCTTGAGATCTTTCAGCTTCGAGCCCT 97782
Qy 121 CTCGGGAGCCCGACGGGCCCTCTGCGGGGCCCGCCAGCGCCCTCTGCGGGGACCGCAAG 180-
Db 97783 CTCGGGAGCCCGACGGGCCCTCTGCGGGGCCCGCCAGCGCCCTCTGCGGGGACCGCAAG 97842
Qy 181 CGCAGCGCGAGGGTCTCTACCTCGAGT----- 209
|||||

```



repeat\_region 3941. .4070  
/rpt\_family="Alu"  
complement(4449. .4724)  
/rpt\_family="Alu"  
complement(5117. .5389)  
/rpt\_family="Alu"  
7668. .7913  
/rpt\_family="Alu"  
complement(7924. .8226)  
/rpt\_family="Alu"  
8250. .8551  
/rpt\_family="Alu"  
complement(8893. .8973)  
/rpt\_family="MIR"  
complement(11829. .12096)  
/rpt\_family="Alu"  
12416. .12740  
/rpt\_family="Alu"  
complement(13181. .13247)  
/rpt\_family="Alu"  
complement(13498. .13647)  
/rpt\_family="LI"  
13869. .14279  
/rpt\_family="Alu"  
18033. .18284  
/rpt\_family="Alu"  
19178. .19451  
/rpt\_family="Alu"  
19458. .19749  
/rpt\_family="Alu"  
20862. .21002  
/rpt\_family="MERS"  
21566. .21832  
/rpt\_family="Alu"  
22231. .22521  
/rpt\_family="Alu"  
complement(22616. .22679)  
/rpt\_family="Alu"  
complement(22869. .23145)  
/rpt\_family="Alu"  
complement(23478. .23772)  
24184. .24372  
/rpt\_family="Alu"  
24529. .24813  
/rpt\_family="Alu"  
complement(25060. .25342)  
/rpt\_family="Alu"  
complement(25367. .25653)  
25787. .26076  
/rpt\_family="Alu"  
26491. .26762  
/rpt\_family="Alu"  
26866. .27143  
/rpt\_family="Alu"  
complement(27620. .27919)  
/rpt\_family="Alu"  
29525. .29719  
/rpt\_family="MIR"  
29762. .30333  
/rpt\_family="Alu"  
30365. .30653  
/rpt\_family="Alu"  
31430. .31722  
/rpt\_family="Alu"  
31765. .31857  
/rpt\_family="Alu"  
32265. .32563  
/rpt\_family="Alu"  
complement(32710. .32858)  
/rpt\_family="Alu"  
complement(33374. .33584)

repeat\_region /rpt\_family="MER2"  
complement(34547. .34595)  
/rpt\_family="MIR"  
repeat\_region complement(36673. .36954)  
/rpt\_family="Alu"  
repeat\_region complement(37002. .37230)  
/rpt\_family="Alu"  
repeat\_region complement(40018. .40170)  
/rpt\_family="Alu"  
variation 40452  
/note="high quality variation versus 3' overlapping clone"  
repeat\_region 40509. .41109  
/rpt\_family="Alu"  
repeat\_region 42609. .42892  
/rpt\_family="Alu"  
repeat\_region 43240. .43370  
/rpt\_family="MIR"  
repeat\_region 43493. .43779  
/rpt\_family="Alu"  
BASE COUNT 11505 a 10560 c 11226 g 10827 t  
ORIGIN  
Query Match 88.0%; Score 1080.4; DB 9; Length 44118;  
Best Local Similarity 91.4%; Pred No. 1e-238;  
Matches 1203; Conservative 0; Mismatches 1; Indels 112; Gaps 1;  
Qy 1 ATGTGTCACCTCTGCAGCTGCCACCCGACCATGACCTCTCGAGGCCCGGACCCGGCC 60  
Db 14601 ATGTGTCACCTCTGCAGCTGCCACCCGACCATGACCTCTCGAGGCCCGGACCCGGCC 14660  
Qy 61 CCCTCCACCATCCCGGGACCCCGGGGGGCTCCGCTCTGAGATCTTACCTTCACCTT 120  
Db 14661 CCCTCCACCATCCCGGGACCCCGGGGGGCTCCGCTCTGAGATCTTACCTTCACCTT 14720  
Qy 121 CTCGGGAGCCGCGAGCGGCCCTCTCGGGGGCCCGACGCTCTCGGGGACCGAAAG 180  
Db 14721 CTCGGGAGCCGCGAGCGGCCCTCTCGGGGGCCCGACGCTCTCGGGGACCGAAAG 14780  
Qy 181 CGCAGCCGAGGGTTCTTACCTTCGAGT----- 209  
Db 14781 CGCAGCCGAGGGTTCTTACCTTCGAGTGTGAGTATCGCGGAAGTGGGCAATCGCGGG 14840  
Qy 210 ----- 209  
Db 14841 GTGCGTCCCTGGAGTCACCTGGGAACGACCCGACTCCAGAGGCTCGACCTGACCTGT 14900  
Qy 210 -----GGTCCGGCGCAGCTGCCAGTCGAGGAAACCGAACCCACCC 248  
Db 14901 CTCCTGTTTTGTCCTCCCTTAGGTCCGGCGCAGCTGCCAGTCGAGGAAACCCACCC 14960  
Qy 249 CAAAGGCTTCTTTCTGCTGCTCACCATCGTCTTTGCCAGATCTTGCATGGCTGAAGA 308  
Db 14961 CAAAGGCTTCTTTCTGCTGCTCACCATCGTCTTTGCCAGATCTTGCATGGCTGAAGA 15020  
Qy 309 GGGTGTGCGGGGCGCCCTGCTCCAGAGAGAGCCGCTTACGCGCGGATCTCTGGCGCCAC 368  
Db 15021 GGGTGTGCGGGGCGCCCTGCTCCAGAGAGAGCCGCTTACGCGCGGATCTCTGGCGCCAC 15080  
Qy 369 CCCTGTGTCCCGGCTCTCGAGCCCTTTAACTGTACTTCGGAGCCCTCGGACTACCGCTCT 428  
Db 15081 CCCTGTGTCCCGGCTCTCGAGCCCTTTAACTGTACTTCGGAGCCCTCGGACTACCGCTCT 15140  
Qy 429 GGACTCAGCATTCTTCTCCAGCAACACCGGCCGCTTCTTAACCTGTGACTTCCCGGCACT 488  
Db 15141 GGACTCAGCATTCTTCTCCAGCAACACCGGCCGCTTCTTAACCTGTGACTTCCCGGCACT 15200  
Qy 489 CCCAAAAAGATCCGAAAAACCAAAACACACCGGCTTACCTGTGCGCGAGAGCG 548  
Db 15201 CCCAAAAAGATCCGAAAAACCAAAACACACCGGCTTACCTGTGCGCGAGAGCG 15260  
Qy 549 TATCCCCAAGTGGGACTTCCGAGGCACTTGAACCTCAGAACTACAGCGGAGACCCAC 608  
|||||

COMMENT	Overlapping Sequences:	Contact: Contact ID:

Db	372	CGGCCACCCCTGTGTCCTCCCGCTCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGAC	431
Qy	421	TACGCTCTGGACCTCAGACACTTCTCCAGCAACACCCGCCCTTCTTAACCTGTGACTC	480
Db	432	TACGCTCTGGACCTCAGACACTTCTCCAGCAACACCCGCCCTTCTTAACCTGTGACTC	491
Qy	481	CCCCACTCCCAAAAGATCCGAAACACCAAAACACACACACACACACACACACACACAC	540
Db	492	CCCCACTCCCAAAAGATCCGAAACACCAAAACACACACACACACACACACACACACAC	551
Qy	541	CGAGCGGTATCCCAACTGGACTCCGAGGCAACTTGAACCTCAGAACACTACAGCGGA	600
Db	552	CGAGCGGTATCCCAACTGGACTCCGAGGCAACTTGAACCTCAGAACACTACAGCGGA	611
Qy	601	GACGCCACCGGTCTTGAAGCGGACCGAGGCGACAGACAGAGCGGCGCATAGAGACC	660
Db	612	CACGCCACCGGTCTTGAAGCGGACCGAGGCGACAGACAGAGCGGCGCATAGAGACC	671
Qy	661	GAGCAGACCCAGCTGGGCTAGGCGGTGGGAAGAGAGCGCTGTTAATTTATTTCT	720
Db	672	GAGCAGACCCAGCTGGGCTAGGCGGTGGGAAGAGAGCGCTGTTAATTTATTTCT	731
Qy	721	TATTGCTCCTAATTAATTTATATGATGATTTATGATGATGATGATGATGATGATGATG	780
Db	732	TATTGCTCCTAATTAATTTATATGATGATTTATGATGATGATGATGATGATGATGATG	791
Qy	781	GTACGTAATATTTTAACTTATGCAAGGCTGTGAGATGTTCCCTCTGCTGTAATATGC	840
Db	792	GTACGTAATATTTTAACTTATGCAAGGCTGTGAGATGTTCCCTCTGCTGTAATATGC	851
Qy	841	AGGCTCTGTGATTTATGACTTTGTGGACTGTTGGAGAGGAGGAGGAGGAGGAGGAGG	900
Db	852	AGGCTCTGTGATTTATGACTTTGTGGACTGTTGGAGAGGAGGAGGAGGAGGAGGAGG	911
Qy	901	GCCAAAGTAGCAGAGAAATGGGAGGACTCGGCTGGGGAGGAGGAGGAGGAGGAGGAGG	960
Db	912	GCCAAAGTAGCAGAGAAATGGGAGGACTCGGCTGGGGAGGAGGAGGAGGAGGAGGAGG	971
Qy	961	AAGTCTGTGTGGTCTAGTTAGAGAGTGACTGATGATGATGATGATGATGATGATGATG	1020
Db	972	AAGTCTGTGTGGTCTAGTTAGAGAGTGACTGATGATGATGATGATGATGATGATGATG	1031
Qy	1021	TCTGTCTACTGTGTAGACTTCGGGACCATTAGGAATGAGATCCGTGAGATCCTTCCA	1080
Db	1032	TCTGTCTACTGTGTAGACTTCGGGACCATTAGGAATGAGATCCGTGAGATCCTTCCA	1091
Qy	1081	TCCTCTTGAAGTCCGCTTTAGGCTGGCTCGAGGTAGAGGCTTGGGGTGGCTGGGCTGT	1140
Db	1092	TCCTCTTGAAGTCCGCTTTAGGCTGGCTCGAGGTAGAGGCTTGGGGTGGCTGGGCTGT	1151
Qy	1141	CACGAGCGACTGTGGAGATCGCCTAGTATGTTCTGTGAACACAAATAAATGATTTAC	1200
Db	1152	CACGAGCGACTGTGGAGATCGCCTAGTATGTTCTGTGAACACAAATAAATGATTTAC	1211
Qy	1201	TCTCAAAAAAAAAAAAA 1219	
Db	1212	TCTCTGCTAAAAA 1230	
RESULT 2			
LOCUS	S81914	1223 bp	mRNA
DEFINITION	IEX-1-radiation-inducible immediate-early gene [human, placenta, mRNA Partial, 1223 nt].		
ACCESSION	S81914		
VERSION	S81914.1	GI:1488384	
KEYWORDS	human placenta.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 1223)		
AUTHORS	Kondratyev,A.D., Chung,K.N. and Jung,M.O.		

TITLE	Identification and characterization of a radiation-inducible glycosylated human early-response gene		
JOURNAL	Cancer Res. 56 (7), 1498-1502 (1996)		
MEDLINE	96181295		
REMARK	GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 177109] from the original journal article. This sequence comes from Fig. 2A.		
FEATURES	Location/Qualifiers		
source	1..1223		
gene	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	19..489		
	/gene="IEX-1"		
CDS	/note="radiation-inducible immediate-early gene"		
	19..489		
	/gene="IEX-1"		
	/note="radiation-inducible immediate-early gene; This sequence comes from Fig. 2A"		
	/codon_start=1		
	/protein_id="AAB36278.1"		
	/db_xref="GI:1488385"		
	/translation="MCHSRSCHPTMTILOATPAPSTIPGRRSGPEITFDPLPEP		
	AAAPAGRPSSGRHRRKRVLYPRVRRQLPVPEPNPAKRI.LFLLLTIVFCQLLMAE		
	EGVRAPLPEDAPNAASLAPTVPVLEPENLTSEPSYALDLSLELQHPAAF"		
BASE COUNT	249 a	368 c	264 t
ORIGIN			
Query Match	93.8%; Score 1152.2; DB #5; Length 1223;		
Best Local Similarity	99.4%; Pred. No. 3.7e-255;		
Matches 1198;	Conservative	0;	Mismatches 3; Indels 4; Gaps 4;
Qy	1	ATGTGTCACTCTCGAGCTGCCAGCTCCACCCGACCACTGACCACTCTCGAGGCCCCGACCCGCGCC	60
Db	19	ATGTGTCACTCTCGAGCTGCCAGCTCCACCCGACCACTGACCACTCTCGAGGCCCCGACCCGCGCC	78
Qy	61	CCCTCCACCATCCGGGACCCCGGGGCTCGGTCTTGAGATCTTACCTTCGACCTT	120
Db	79	CCCTCCACCATCCGGGACCCCGGGGCTCGGTCTTGAGATCTTACCTTCGACCTT	138
Qy	121	CTCCGAGGCCCGACGGCCCTTCGCCGGCGCCCGCCAGCCCTTCGCCGGGACCCGAAAG	180
Db	139	CTCCGAGGCCCGACGGCCCTTCGCCGGCGCCCGCCAGCCCTTCGCCGGGACCCGAAAG	198
Qy	181	CGAGCCGCGAGGTTCTCTACCTTCGAGTGGTCCGGCGCCAGCTGCCAGTCGAGGACCG	240
Db	199	CGAGCCGCGAGGTTCTCTACCTTCGAGTGGTCCGGCGCCAGCTGCCAGTCGAGGACCG	258
Qy	241	AACCCAGCCAAAAGGCTTCTTTCTGCTCACTACCATCGTCTTCTTCTCCAGATCTCTGATG	300
Db	259	AACCCAGCCAAAAGGCTTCTTTCTGCTCACTACCATCGTCTTCTTCTCCAGATCTCTGATG	318
Qy	301	GCTGAAGAGGTTGTGCGGCGCCCTTCCTCCAGAGGACGCCCTTAACGCCGATCCCTG	360
Db	319	GCTGAAGAGGTTGTGCGGCGCCCTTCCTCCAGAGGACGCCCTTAACGCCGATCCCTG	378
Qy	361	CGCCCAACCTCTGTGTCGCCGCTCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGAC	420
Db	379	CGCCCAACCTCTGTGTCGCCGCTCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGAC	438
Qy	421	TACGCTCTGAGCTTCAGCACTTTCCTCCAGCAACACCCGCCCTTCTTAACCTGTGACTC	480
Db	439	TACGCTCTGAGCTTCAGCACTTTCCTCCAGCAACACCCGCCCTTCTTAACCTGTGACTC	498
Qy	481	CCCGACTTCCCAAAAAGAAATCCGAAAACACCAAGAAACACACAGCGGTACCTGGTGG	540
Db	499	CCCGACTTCCCAAAAAGAAATCCGAAAACACCAAGAAACACACAGCGGTACCTGGTGG	558
Qy	541	CGAGAGCGTATCCCAACTTGGGACTTCGAGGCGAATTTGAACCTCAGAACACTACAGCGGA	600
Db	559	CGAGAGCGTATCCCAACTTGGGACTTCGAGGCGAATTTGAACCTCAGAACACTACAGCGGA	618
Qy	601	GACGCCACCCGCTGCTTGGAGCGGACCGAGGCCACAGACCGAGGCGCATAGAGACC	660



c 22 64 5.2 209887 71 AC079424 Mus muscu  
23 63.4 5.2 64739 62 AC022717 Homo sapi  
c 24 62.6 5.1 300695 71 AC079431 Mus muscu  
c 25 60.8 5.0 187738 66 AC027682 Homo sapi  
26 59.6 4.9 134940 60 AC018939 Homo sapi  
27 59.4 4.8 50602 56 AC011100 Homo sapi  
c 28 59.4 4.8 120379 56 AC011429 Homo sapi  
c 29 59.4 4.8 220469 70 AC074307 Mus muscu  
c 30 59.2 4.8 62649 62 AC022552 Homo sapi  
31 58.8 4.8 78765 61 AC021584 Homo sapi  
32 58.6 4.8 259451 58 AC015605 Homo sapi  
33 58.4 4.8 149915 61 AC021637 Homo sapi  
34 58.2 4.7 75628 61 AC021793 Homo sapi  
c 35 58.2 4.7 135217 71 AC079081 Homo sapi  
36 58 4.7 79023 61 AC021951 Homo sapi  
37 58 4.7 87440 61 AC021226 Homo sapi  
38 58 4.7 150695 63 AC023576 Homo sapi  
c 39 57.8 4.7 200580 72 AC084064 Homo sapi  
40 57.6 4.7 141892 62 AC023197 Homo sapi  
41 57.4 4.7 76795 63 AC023515 Homo sapi  
42 57.2 4.7 840 7 CNS01BNS Botrytis  
43 57.2 4.7 73948 62 AC022556 Homo sapi  
c 44 57.2 4.7 258174 71 AC079429 Mus muscu  
c 45 57 4.6 249262 71 AC079430 Mus muscu

ALIGNMENTS

RESULT 1  
HSDIF2 HSDIF2 1230 bp mRNA PRI 12-AUG-1997  
LOCUS Homo sapiens mRNA for DIF-2 protein.  
DEFINITION Y14551  
ACCESSION Y14551.1 GI:2660550  
VERSION dif-2 gene.  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1230)  
AUTHORS Pietzsch,A., Buchler,C., Aslanidis,C. and Schmitz,G.  
TITLE Identification and characterization of a novel monocyte/macrophage  
differentiation-dependent gene that is responsive to  
lipopolysaccharide, ceramide, and lysophosphatidylcholine  
JOURNAL Biochem. Biophys. Res. Commun. 235 (1), 4-9 (1997)  
MEDLINE 97339426  
REFERENCE 2 (bases 1 to 1230)  
AUTHORS Pietzsch,A.  
TITLE Direct Submission  
JOURNAL Submitted (11-AUG-1997) A. Pietzsch, Institute for Clinical  
Chemistry and Laboratory Medicine, University of Regensburg,  
Franz-Josef-Strauss-Allee 11, 93053 Regensburg, FRG  
REFERENCE 3 (bases 1 to 1230)  
AUTHORS Kondratyev,A.D., Chung,K.N. and Jung,M.O.  
TITLE Identification and characterization of a radiation-inducible  
glycosylated human early-response gene  
JOURNAL Cancer Res. 56 (7), 1498-1502 (1996)  
MEDLINE 96181295  
FEATURES  
source Location/Qualifiers  
1..1230  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="leukocytes"  
c1..11  
/gene="dif-2"  
1..1230  
/gene="dif-2"  
12..482  
/gene="dif-2"  
/codon\_start=1  
/product="DIF-2 protein"  
/protein\_id="CAA74886.1"

/db\_xref="GI:2660551"  
/db\_xref="Swiss-Prot:P46695"  
/translation="MCHSRSCHTMTILQAPTPAPSTIPGRRSGPELTFDPLPER  
AAPAGRPASRGHRKSRRLVPRVRRQLPVEPNPARKRLLELLLTIVFCOTIMAE  
EGVPAPLPEDAPNAASLAPTPVSPVLEPFLNLTSEPSDYALDLSTFLQOHPAAP"  
172  
conflict /gene="dif-2"  
/citation=[3]  
/replace="g"  
328  
conflict /gene="dif-2"  
/citation=[3]  
/replace="g"  
483..>1230  
/gene="dif-2"  
673..675  
conflict /gene="dif-2"  
/citation=[3]  
/replace="ag"  
687..689  
conflict /gene="dif-2"  
/citation=[3]  
/replace="tg"  
783..785  
conflict /gene="dif-2"  
/citation=[3]  
/replace="tg"  
1020..1021  
conflict /gene="dif-2"  
/citation=[3]  
/replace="att"  
1120  
conflict /gene="dif-2"  
/citation=[3]  
/replace="g"  
1133  
conflict /gene="dif-2"  
/citation=[3]  
/replace="t"  
BASE COUNT 259 a 367 c 342 g 262 t  
ORIGIN  
Query Match 98.4%; Score 1207.8; DB 53; Length 1230;  
Best Local Similarity 99.4%; Pred. No. 5.7e-268;  
Matches 1212; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Qy 1 ATGTGTACTCTCGAGCTGCCACCGGAGCATGACATCTCGAGCCCGGCGCCG 60  
Db 12 ATGTGTACTCTCGAGCTGCCACCGGAGCATGACATCTCGAGCCCGGCGCCG 71  
Qy 61 CCTCCACCATCCGGGACCGGCGGCTCCGGTCTGAGATCTTACCTTCGACCT 120  
Db 72 CCTCCACCATCCGGGACCGGCGGCTCCGGTCTGAGATCTTACCTTCGACCT 131  
Qy 121 CTCGGGAGCCGCGAGCGCCCTGCGGGCGCCCGAGCGCCCTCTCGGGGACCGA 180  
Db 132 CTCGGGAGCCGCGAGCGCCCTGCGGGCGCCCGAGCGCCCTCTCGGGGACCGA 191  
Qy 181 CGCAGCGGAGGGTCTCTACCTCGAGTGTGCGGGCGGCGAGCTGCGAGTCGAG 240  
Db 192 CGCAGCGGAGGGTCTCTACCTCGAGTGTGCGGGCGGCGAGCTGCGAGTCGAG 251  
Qy 241 AACCCAGCAAAAGGCTTCTCTTCTGCTGCTCACCATCTTCTGCGCAGATCCTG 300  
Db 252 AACCCAGCAAAAGGCTTCTCTTCTGCTGCTCACCATCTTCTGCGCAGATCCTG 311  
Qy 301 GCTGAAGAGGGTGTGCGGGCGCCCTGCGCTCCAGAGGAGCGCCCTAACCCGCA 360  
Db 312 GCTGAAGAGGGTGTGCGGGCGCCCTGCGCTCCAGAGGAGCGCCCTAACCCGCA 371  
Qy 361 GCGCCGACCCCTGTGTCCCGCCCTCTCGAGCCCTTAACTGACTTCGAGGCCCTCG 420

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 12:11:54 ; Search time 1900.13 Seconds  
(without alignments)  
3307.456 Million cell updates/sec

Title: US-08-799-910-9  
Perfect score: 1228  
Sequence: 1 ATGCTCACTCTCGACGCTG.....AAAAAAAAAAAACTCGAG 1228

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues 2236266  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba1:  
2: gb\_ba2:  
3: gb\_om:  
4: gb\_ov:  
5: gb\_ph:  
6: gb\_pl1:  
7: gb\_pl2:  
8: gb\_pr1:  
9: gb\_pr2:  
10: gb\_pr3:  
11: gb\_ro:  
12: gb\_sy:  
13: gb\_un:  
14: em\_fun:  
15: em\_hum1:  
16: em\_hum2:  
17: em\_in:  
18: em\_om:  
19: em\_or:  
20: em\_ov:  
21: em\_pat:  
22: em\_ph:  
23: em\_pi:  
24: em\_ro:  
25: em\_sts:  
26: em\_sy:  
27: em\_un:  
28: em\_vi:  
29: gb\_ba3:  
30: gb\_in1:  
31: gb\_in2:  
32: gb\_in3:  
33: gb\_pl3:  
34: gb\_pr4:  
35: em\_ba1:  
36: em\_bt1:  
37: em\_bt2:  
38: em\_bt3:  
39: em\_bt4:  
40: em\_bt5:  
41: em\_bt6:  
42: em\_bt7:  
43: em\_bt8:

44: em\_htg8:  
45: em\_htg9:  
46: em\_htg10:  
47: em\_hum3:  
48: em\_hum4:  
49: em\_hum5:  
50: em\_hum6:  
51: gb\_pr5:  
52: gb\_pr6:  
53: gb\_pr7:  
54: gb\_htg1:  
55: gb\_htg2:  
56: gb\_htg3:  
57: gb\_htg4:  
58: gb\_htg5:  
59: gb\_htg6:  
60: gb\_htg7:  
61: gb\_htg8:  
62: gb\_htg9:  
63: gb\_htg10:  
64: gb\_htg11:  
65: gb\_htg12:  
66: gb\_htg13:  
67: gb\_htg14:  
68: gb\_htg15:  
69: gb\_htg16:  
70: gb\_htg17:  
71: gb\_htg18:  
72: gb\_htg19:  
73: gb\_htg20:  
74: gb\_htg21:  
75: gb\_htg22:  
76: gb\_htg23:  
77: gb\_sts1:  
78: gb\_sts2:  
79: gb\_vil:  
80: gb\_vil2:  
81: gb\_pat1:  
82: gb\_pat2:  
83: em\_htg0:  
84: gb\_htg24:  
85: gb\_pr8:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		DB ID	Description
		Match	Length		
1	1207.8	98.4	1230	53	Y14551 Homo sapien
2	1152.2	93.8	1223	85	S81914 IEX-1-radia
3	1080.4	88.0	44118	9	AC006165 Homo sapi
4	1080.4	88.0	90244	8	AB023051 Homo sapi
5	1080.4	88.0	200000	51	AP000512 Homo sapi
6	1048.6	85.4	1309	10	AF039067 Homo sapi
7	1048.6	85.4	1693	10	AF071596 Homo sapi
8	979.8	79.8	1864	53	X96438 H.sapiens p
9	471	38.4	477	10	AF083421 Homo sapi
10	298.8	24.3	343	52	AJ227914 Homo sapi
11	242.4	19.7	1758	11	X96437 R.norvegicu
12	210.6	17.1	203476	70	AC074150 Mus muscu
13	209	17.0	1938	11	MMGLY96
14	202	16.4	187868	62	AC022301 Mus muscu
15	151	12.3	297	81	A75448 Sequence 11
16	151	12.3	297	81	A78427 Sequence 11
17	122	9.9	122	78	G62116 1899 Human
18	74	6.0	205512	74	AL355388 Homo sapi
19	70.4	5.7	267	81	A75468 Sequence 11
20	70.4	5.7	267	81	A78447 Sequence 11
21	64	5.2	83276	65	AC026631 Homo sapi

	Query Match	15.4%	Score 68:	DB 1;	Length 1395;
	Best Local Similarity	43.2%	Pred. NO. 36;		
	Matches	19:	Conservative	6:	Mismatches 13; Indels
				6:	Gaps
Oy	28	ILMAEEGVAP- -LPPEDAPNAASLAPTVPSPVLEPFLNTSEPS	69		
		:   :   :	:   :   :		
		:   :   :	:   :   :		
db	44	IADRPGLGPEHSPSSQDSSPSPTSP-SPVLE- --GSEP	83		
		:   :   :	:   :   :		

Search completed: January 30, 2001, 23:12:25  
Job time: 4006 sec

[illegible]







RA Angrand P.O., Valvatne H., Jeannougin F., Adamson A.,  
RA van der Hoeven F., Olsen L., Tekotte H., Huang N., Poch O.,  
RA Lamerdin J., Chambon P., Losson R., Stewart A., Aasland R.;  
RT "Mammalian trithorax" and ASH1-like proteins: putative chromatin  
RT regulators which contain PHD fingers and SET domains.";  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RN SEQUENCE FROM N.A. (LONG ISOFORM).  
RP Lamerdin J.E., McCreedy P.M., Adamson A.W., Burkhardt-Schultz K.,  
RA Garcia E., Kyle A., Ramirez M., Stillwagen S., Garges J., Danganan L.,  
RA Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,  
RA Carrano A.V.;  
RT "Sequence analysis of a 1 Mb region in human 19q13.1";  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RN SEQUENCE OF 816-2715 FROM "N.A. (LONG ISOFORM).  
RP TISSUE-BRAIN;  
RX MEDLINE=97349984; PubMed=9205841;  
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. VII.  
RT The complete sequences of 100 new cDNA clones from brain which can  
RT code for large proteins in vitro.";  
RL DNA Res. 4:141-150(1997).  
RN [4]  
RN SEQUENCE OF 111-2715 FROM N.A. (LONG ISOFORM).  
RP TISSUE-TESTIS, AND LEUKOCYTE;  
RX MEDLINE=20105772; PubMed=10637508;  
RA Huntsman D.G., Chin S.-F., Muleris M., Batley S.J., Collins V.P.,  
RA Wiedemann L.M., Aparicio S., Caldas C.;  
RT "MLL2, the second human homolog of the Drosophila trithorax gene, maps  
RT to 19q13.1 and is amplified in solid tumor cell lines.";  
RL Oncogene 18:7975-7984(1999).  
RN [5]  
RN PARTIAL SEQUENCE FROM N.A. (LONG AND TRUNCATED ISOFORMS).  
RP TISSUE-PLACENTA, AND BONE MARROW;  
RX MEDLINE=99339983; PubMed=10409430;  
RA FitzGerald K.T., Diaz M.O.;  
RT "MLL2: A new mammalian member of the trx/MLL family of genes.";  
RL Genomics 59:187-192(1999).  
CC -!- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -!- TRUNCATED FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST LEVELS IN TESTIS.  
CC ALSO FOUND IN BRAIN, BONE MARROW, HEART, MUSCLE, KIDNEY, PANCREAS,  
CC SPLEEN, THYMUS, PROSTATE, OVARY, INTESTINE, COLON, PERIPHERAL  
CC BLOOD LYMPHOCYTES, AND PLACENTA.  
CC -!- DISEASE: OFTEN AMPLIFIED IN PANCREATIC CARCINOMAS.  
CC -!- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION  
CC FACTORS.  
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 'SET' DOMAIN.  
CC -!- SIMILARITY: CONTAINS 3 PHD-FINGER DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AJ007041; CAB45385.1; -;  
DR EMBL: AD000671; -; NOT\_ANNOTATED\_CDS.  
DR EMBL: AB002302; BAA20763.2; -;  
DR EMBL: AF186605; AAD56420.1; -;  
DR EMBL: AF104918; AAD17932.1; -;  
DR EMBL: AF105279; AAD26113.1; -;  
DR EMBL: AF105280; AAD26112.1; -;  
DR INTERPRO: IPR001214; -;  
DR INTERPRO: IPR001965; -;  
DR INTERPRO: IPR002857; -;

DR PFAM: PF00628; PHD; 3.  
DR PFAM: PF00856; SET; 1.  
DR PFAM: PF02008; zf-CXXC; 1.  
KW DNA-binding; Bromodomain; Nuclear protein; Zinc-finger; Metal-binding;  
KW Transcription regulation; Alternative splicing.  
FT DNA\_BIND 37 44 A..T HOOK (BY SIMILARITY).  
FT DNA\_BIND 110 117 A..T HOOK (BY SIMILARITY).  
FT DNA\_BIND 120 127 A..T HOOK (BY SIMILARITY).  
FT DOMAIN 1203 1252 PHD 1.  
FT DOMAIN 1253 1303 PHD 2.  
FT DOMAIN 1337 1396 PHD 3.  
FT DOMAIN 1449 1471 BROMODOMAIN (DIVERGENT).  
FT DOMAIN 2586 2715 SET DOMAIN.  
FT DOMAIN 26 37 POLY-GLY.  
FT DOMAIN 248 255 POLY-PRO.  
FT DOMAIN 362 398 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 402 771 PRO-RICH.  
FT DOMAIN 808 812 POLY-GLN.  
FT DOMAIN 1963 1970 POLY-PRO.  
FT DOMAIN 2251 2259 POLY-PRO.  
FT VARSPPLIC 532 582 VSARSRVKTPRRRNDEDPKPKPKVSPVLRPPITTSPP  
FT VPOEPAPVPS -> PLUSQLLPMTLQLSLGQWAAPTTS  
FT ACIDSLPLMSLLRPRCLPTGLQL (IN ISOFORM  
FT TRUNCATED).  
FT VARSPLIC 583 2715 MISSING (IN ISOFORM TRUNCATED).  
FT CONFLICT 834 834 K -> E (IN REF. 5).  
FT CONFLICT 941 941 S -> Y (IN REF. 5).  
FT CONFLICT 1317 1317 H -> Q (IN REF. 5).  
FT CONFLICT 1362 1362 E -> Y (IN REF. 5).  
FT CONFLICT 1438 1438 D -> N (IN REF. 5).  
FT CONFLICT 2622 2622 D -> H (IN REF. 5).  
SQ SEQUENCE 2715 AA; 293511 MW; C0615B981BBB7BF CRC64;  
Query Match 16.0%; Score 70.5; DB 1; Length 2715;  
Best Local Similarity 34.5%; Pred. No. 43;  
Matches 19; Conservative 8; Mismatches 27; Indels 1; Gaps 1;  
QY 8 EEPNPAKRLLFLLLTTFVFCOILMAEEGVAPLPEDAPNNAASLAPT-PVSPVLEP 61  
Db 2182 EPPKATSKILLVKNKGQVFVKMAGEGEPVPPVKQPLPTTISAPTTSWTLP 2236  
RESULT 9  
FAS2\_YEAST STANDARD; PRT: 1894 AA.  
AC P19097;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE FATTY ACID SYNTHASE, SUBUNIT ALPHA (EC 2.3.1.86) [INCLUDES:  
DE EC 1.1.1.100; EC 2.3.1.41].  
GN FAS2 OR YPL231W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomyces.  
RN [1]  
RX MEDLINE=88315020; PubMed=2900835;  
RA Mohamed A.H., Chirala S.S., Mody N.H., Huang W.Y., Wakil S.J.;  
RT "Primary structure of the multifunctional alpha subunit protein of  
RT yeast fatty acid synthase derived from FAS2 gene sequence.";  
RL J. Biol. Chem. 263:12315-12325(1988).  
RN [2]  
RN MUTAGENESIS OF GLY-1257.  
RP STRAIN=S288C;  
RX MEDLINE=94316198; PubMed=8041367;  
RA Inokoshi J., Tomoda H., Hashimoto H., Watanabe A., Takeshima H.,  
RA Omura S.;  
RT "Cerulenin-resistant mutants of Saccharomyces cerevisiae with an  
RT altered fatty acid synthase gene.";  
RL Mol. Gen. Genet. 244:90-96(1994).  
CC -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF









CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; S81914; AAR36278.1; -;  
DR EMBL; X96438; CAA65304.1; -;  
DR EMBL; Y14551; CAA74886.1; -;  
DR EMBL; T49531; -; NOT\_ANNOTATED\_CDS.  
DR MIM; 602996; -;  
KW Glycoprotein; Transmembrane; Signal-anchor  
FT DOMAIN 1 82  
FT TRANSMEM 83 99  
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL).  
FT DOMAIN 100 156 \* EXTRACELLULAR (POTENTIAL).  
FT CARBOHYD 133 133 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CONFLICT 54 54 A -> G (IN REF. 1).  
FT CONFLICT 106 106 P -> R (IN REF. 1).  
FT SEQUENCE 156 AA; 16928 MW; 83C06116C80B9240 CRC64;

Query Match 100.0%; Score 442; DB 1; Length 156;  
Best Local Similarity 100.0%; Pred. No. 3.6e-37;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLPVEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASLAPTPVSPVLE 60  
DB 71 VRRQLPVEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASLAPTPVSPVLE 130

QY 61 PNLTSSEPSDYALDLSTFLQHPAAAF 86  
DB 131 PNLTSSEPSDYALDLSTFLQHPAAAF 156

RESULT 2  
ID IEX1\_MOUSE STANDARD; PRT: 153 AA.  
AC P46694;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY  
DE PROTEIN GLY96) (IMMEDIATE EARLY RESPONSE 3 PROTEIN).  
GN IER3 OR IEX1 OR GLY96.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C;  
RX MEDLINE=93173526; PubMed=8437864;  
RA Charles C.H., Yoon J.K., Simske J.S., Lau L.F.;  
RT "Genomic structure, cDNA sequence, and expression of gly96, a growth  
RT factor-inducible immediate-early gene encoding a short-lived  
RT glycosylated protein.";  
RL Oncogene 8:1797-801(1993).  
CC -!- FUNCTION: NOT KNOWN; EXPRESSED DURING THE G0-G1 TRANSITION OF THE  
CC CELL CYCLE.  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).  
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE LUNG, TESTES  
CC AND THE UTERUS.  
CC -!- INDUCTION: BY SERUM GROWTH FACTORS AND TPA.  
CC -!- PTM: GLYCOSYLATED.  
CC -!- SIMILARITY: STRONG, TO HUMAN ORTHOLOG.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X67644; -; NOT\_ANNOTATED\_CDS.  
DR PIR; S33363; S33363.  
DR MGD; MGI:104814; IER3.  
KW Glycoprotein; Transmembrane; Signal-anchor  
FT DOMAIN 1 85  
FT TRANSMEM 86 102  
FT SIGNAL-ANCHOR (POTENTIAL).  
FT (POTENTIAL).  
FT DOMAIN 103 153  
FT CARBOHYD 137 137 EXTRACELLULAR (POTENTIAL).  
FT SEQUENCE 153 AA; 16875 MW; F4429E3B0120DE11 CRC64;

Query Match 61.8%; Score 273; DB 1; Length 153;  
Best Local Similarity 68.2%; Pred. No. 1.8e-20;  
Matches 58; Conservative 5; Mismatches 14; Indels 8; Gaps 2;

QY 1 VRRQLPVEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASLAPTPVSPVLE 56  
DB 71 VRRQLPVEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASLAPTPVSPVLE 128

QY 57 -PVLEPNLTSEPSDYALDLSTFL 79  
DB 129 APPVLEPLNLTSSEPSDYALDLKAF 153

RESULT 3  
ID SOS\_DROME STANDARD; PRT: 1595 AA.  
AC P28675;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE SON OF SEVENLESS PROTEIN.  
GN SOS  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-OREGON-R;  
RX MEDLINE=92141820; PubMed=1736363;  
RA Bonfini L., Karlovich C.A., Dasgupta C., Banerjee U.;  
RT "The Son of sevenless gene product: a putative activator of Ras.";  
RL Science 255:603-606(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92034991; PubMed=1934068;  
RA Simon M.A., Bowtell D.D.L., Dodson G.S., Lavery T.R., Rubin G.M.;  
RT "Ras1 and a putative guanine nucleotide exchange factor perform  
RT crucial steps in signaling by the sevenless protein tyrosine  
RT kinase.";  
RL Cell 67:701-716(1991).  
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. SOS IS  
CC IMPLICATED IN NEURONAL DEVELOPMENT.  
CC -!- SUBUNIT: MAY FORM A COMPLEX WITH SEVENLESS AND DRK.  
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M83931; AAB04680.1; -;  
DR EMBL; M77501; AAA28904.1; -;  
DR PIR; A41216; A41216.  
DR HSPSP; P04002; 1WFA.

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: January 30, 2001, 23:12:24 ; Search time 43.61 seconds  
(without alignments)  
63.685 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156

Perfect score: 442  
Sequence: 1 VRRQLPVEEPNPAKRLLEL.....EPSDYALDLSTFLQQHPAFA 86

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	442	100.0	156	1	IEXL_HUMAN
2	273	61.8	153	1	IEXL_MOUSE
3	78.5	17.8	1595	1	SOS-DROME
4	73.5	16.6	272	1	GSPC_ERWCH
5	72.5	16.4	272	1	GSQC_ERWCH
6	71.5	16.2	626	1	GPBA_HUMAN
7	70.5	16.0	605	1	BRL1_EBV
8	70.5	16.0	2715	1	TRX2_HUMAN
9	69.5	15.7	1894	1	FAS2_YEAST
10	69	15.6	1402	1	IF4G_RABIT
11	68.5	15.5	181	1	C21Q_HUMAN
12	68.5	15.5	882	1	CAD1_HUMAN
13	68.5	15.5	1337	1	PTPB_HUMAN
14	68.5	15.5	1844	1	POLR_TYMV
15	68	15.4	1395	1	IF4G_HUMAN
16	67	15.2	283	1	EXTN_SORBI
17	67	15.2	586	1	LREL_YEAST
18	67	15.2	4289	1	TENX_HUMAN
19	66.5	15.0	3149	1	TEGU_EBV
20	66	14.9	621	1	ILVB_MYCAV
21	65.5	14.8	325	1	MODD_MYCBO
22	65.5	14.8	325	1	MODD_MYCTO
23	65.5	14.8	329	1	EXBB_PSEPU
24	65.5	14.8	812	1	EBN3_EBV
25	65.5	14.8	1093	1	AF17_HUMAN
26	65	14.7	209	1	CXX1_HUMAN
27	65	14.7	328	1	ZIFA_ECOLI
28	65	14.7	427	1	TEF3_HUMAN
29	65	14.7	724	1	P85B_BOVIN
30	64.5	14.6	686	1	MYBB_CHICK
31	64.5	14.6	1183	1	DRFL_SOYB
32	64	14.5	510	1	C933_SOYBN
33	64	14.5	510	1	IA1_HUMAN

34	64	14.5	537	1	PTNS_HUMAN
35	64	14.5	541	1	PTNS_MOUSE
36	64	14.5	590	1	PTZ1_RHIME
37	64	14.5	966	1	SSN6_YEAST
38	63.5	14.4	443	1	HXA3_HUMAN
39	63.5	14.4	540	1	NUSA_MYCPN
40	63.5	14.4	816	1	GAIF_NEUCR
41	63.5	14.4	1776	1	POLR_OYMV
42	63.5	14.4	1844	1	POLR_TYMV
43	63	14.3	323	1	PF27_MOUSE
44	63	14.3	420	1	EF1A_HALMA
45	63	14.3	598	1	NR41_HUMAN

ALIGNMENTS

RESULT 1					
IEXL_HUMAN					
ID	IEXL_HUMAN	STANDARD:	PRT:	156 AA.	
AC	P46695; Q93044; Q92691;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY				
DE	PROTEIN GLY96) (IMMEDIATE EARLY RESPONSE 3 PROTEIN) (PACAP-RESPONSIVE				
DE	GENE 1 PROTEIN) (PRG1 PROTEIN) (DIFFERENTIATION-DEPENDENT GENE 2				
DE	PROTEIN) (DIF-2 PROTEIN).				
GN	IER3 OR IEX1 OR PRG1 OR DIF2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=PLACENTA;				
RX	MEDLINE=96181295; PubMed=8603392;				
RA	Kondratyev A.D., Chung K.-N., Jung M.O.;				
RT	"Identification and characterization of a radiation-inducible				
RT	glycosylated human early-response gene.";				
RL	Cancer Res. 56:1498-1502(1996).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=96221139; PubMed=8653710;				
RA	Schaefer H., Trausold A., Siegel E.G., Folsch U.R., Schmidt W.E.;				
RT	"PRG1: a novel early-response gene transcriptionally induced by				
RT	pituitary adenylate cyclase activating polypeptide in a pancreatic				
RT	carcinoma cell line.";				
RL	Cancer Res. 56:2641-2641(1996).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97339426; PubMed=9196025;				
RA	Pietzsch A., Buechler C., Aslanidis C., Schmitz G.;				
RT	"Identification and characterization of a novel monocyte/macrophage				
RT	differentiation-dependent gene that is responsive to				
RT	lipopolysaccharide, ceramide, and lysophosphatidylcholine.";				
RL	Biochem. Biophys. Res. Commun. 235:4-9(1997).				
RN	[4]				
RP	PRELIMINARY SEQUENCE OF 1-106 FROM N.A.				
RC	TISSUE=PLACENTA;				
RA	Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman M.,				
RA	Hultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J.,				
RA	Rifkin L., Rohlfing T., Tan F., Trevaskis E., Waterston R.,				
RA	Williamson A., Wohldmann P., Wilson R.;				
RL	Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.				
CC	-!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).				
CC	-!- INDUCTION: BY RADIATION, TPA, OKADAIC ACID AND TNF-ALPHA.				
CC	-!- PTM: GLYCOSYLATED.				
CC	-!- SIMILARITY: STRONG, TO MOUSE ORTHOLOG.				
CC	-!- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO				
CC	FRAMESHIFTS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				



RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman C.S., Pan S., Poillard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of *Drosophila melanogaster*,"  
RL Science 287:2185-2195(2000).  
[5]  
RP SEQUENCE OF 281-349 FROM N.A.  
RX MEDLINE-93066327; PubMed-1438276;  
RA Sommer R.J., Retzlaff M., Goerlich K., Sander K., Tautz D.,  
RT "Evolutionary conservation pattern of zinc-finger domains of  
RT *Drosophila* segmentation genes,"  
RL Proc. Natl. Acad. Sci. U.S.A. 89:10782-10786(1992).  
[6]  
RP POLYMORPHISM.  
RA Tautz D., Nigro L.,  
RT "Microevolutionary divergence pattern of the segmentation gene  
RT hunchback in *Drosophila*,"  
RL Mol. Biol. Evol. 15:1403-1411(1998).  
CC -!- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT  
CC OF HEAD STRUCTURES.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- TISSUE SPECIFICITY: IN EMBRYO, EXPRESSION OF MATERNAL TRANSCRIPT  
CC IS HIGHEST IN ANTERIOR REGION. ZYGOTIC TRANSCRIPT IS EXPRESSED IN  
CC ANTERIOR REGION UNTIL THE BEGINNING OF GASTRULATION AND IN  
CC POSTERIOR REGION UNTIL EARLY GASTRULATION. AFTER THIS, IT IS  
CC EXPRESSED IN DEVELOPING NERVOUS SYSTEM.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED MATERNALLY AND ZYGOTICALLY.  
CC EXPRESSION OF THE MATERNAL TRANSCRIPT DECREASES UNTIL EMBRYONIC  
CC STAGE 14, ZYGOTIC TRANSCRIPT IS FIRST DETECTED AT STAGE 11.  
CC -!- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: Y00274; CAA68377.1; -  
DR EMBL: U17742; AAB60232.1; -  
DR EMBL: AE003680; AAF54270.1; -  
DR PIR: A29253; A29253.  
DR TRANSFAC: T00395; -  
DR FLYBASE: FBgn0001180; hb.  
DR INTERPRO: IPR000822; -  
DR PFAM: PF00096; zf-C2H2; 6.  
DR PRINTS: PR00048; ZINC\_FINGER.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2; 3.  
KW Developmental protein; Gap; zinc-finger; Metal-binding; DNA-binding;  
KW Repeat; Nuclear protein; Polymorphism.  
FT DOMAIN 240 349  
FT ZINC FINGERS I.  
FT DOMAIN 705 757  
FT ZINC FINGERS II.  
FT DOMAIN 61 69  
FT POLY-GLN.  
FT DOMAIN 102 108  
FT POLY-GLN.  
FT DOMAIN 117 122  
FT POLY-HIS.  
FT ZN\_FING 240 262  
FT C2H2-TYPE.  
FT ZN\_FING 269 291  
FT C2H2-TYPE.  
FT ZN\_FING 297 319  
FT C2H2-TYPE.  
FT ZN\_FING 325 349  
FT C2H2-TYPE.  
FT DOMAIN 404 409  
FT POLY-GLN.  
FT DOMAIN 555 561  
FT POLY-GLN.  
FT ZN\_FING 705 727  
FT C2H2-TYPE.

FT ZN\_FING 733 757 C2H2-TYPE.  
FT VARIANT 437 437 A -> P.  
FT VARIANT 649 649 V -> M.  
SQ SEQUENCE 758 AA; 83113 MW; AD74802EBB56ACD7 CRC64;

Query Match 11.5%; Score 94.5; DB 1; Length 758;  
Best Local Similarity 24.6%; Pred. No. 1.5;  
Matches 43; Conservative 19; Mismatches 56; Indels 57; Gaps 7;

OY 2 CHSRSCH-----PTMTILOAPTTPAPS---TIPGPRRGSGPEIFFDPLPEPAAPAGR 51  
DB 337 CHSFKLHLRKYGKPGMVLDEGTPNPSLVIDVYGRRG-----PKSKNGG 382  
OY 52 PSASRCHRRKRSRVLYPRVVRQLPVEEPNPAKRLLFLLLTIVFCQILMAEGVP----- 106  
DB 383 PIASGSGSGSRKSNVAAPQO---QOSOPAQ-----PVATLSAALOGFPLVQON 432  
OY 107 -----APLPEDAPNAASLAPTVPVSPVLEPNLTSEPSDYALDLSTFLOQH 152  
DB 433 SAPPAASPVLPASPASPAKSVASVEQTPSLP-----SPANLPPPLASLQON 478

Search completed: January 30, 2001, 23:12:24  
Job time: 4005 sec





RT "Molecular cloning of chicken myosin-binding protein (MyBP) H (86-kDa  
RT protein) reveals extensive homology with MyBP-C (C-protein) with  
RT conserved immunoglobulin C2 and fibronectin type III motifs."  
RL J. Biol. Chem. 268:3670-3676(1993).  
CC -I- FUNCTION: BINDS TO MYOSIN; PROBABLY INVOLVED IN INTERACTION WITH  
CC THICK MYOFILAMENTS IN THE A-BAND.  
CC -I- TISSUE SPECIFICITY: SKELETAL MUSCLE. SEEMS TO BE ALSO EXPRESSED IN  
CC THE SLOW TONIC AID MUSCLE. NOT DETECTED IN GIZZARD OR HEART.  
CC -I- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
CC -I- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
CC -I- SIMILARITY: BELONGS TO THE MYBP FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L05605; AAA21418.1; .  
DR INTERPRO: IPR001777; .  
DR INTERPRO: IPR002965; .  
DR INTERPRO: IPR003006; .  
DR PFAM; PF00041; fn3; 2.  
DR PFAM; PF00047; ig; 2.  
DR PRINTS; PR00014; FNTYPEIII.  
DR PRINTS; PR01217; PRICHEXTEN.  
KW Immunoglobulin domain; Cell adhesion; Muscle protein; Thick filament;  
KW Repeat.  
FT DOMAIN 135 221 FIBRONECTIN TYPE-III.  
FT DOMAIN 253 312 IG-LIKE C2-TYPE DOMAIN.  
FT DOMAIN 331 416 FIBRONECTIN TYPE-III.  
FT DOMAIN 458 518 IG-LIKE C2-TYPE DOMAIN.  
FT CONFLICT 2 2 T -> G (IN AA SEQUENCE).  
FT CONFLICT 9 9 A -> P (IN AA SEQUENCE).  
FT CONFLICT 15 15 A -> K (IN AA SEQUENCE).  
FT SEQUENCE 537 AA; 58678 MW; 06C4CF0EF1DD233 CRC64;  
-----  
Query Match 11.6%; Score 95.5; DB 1; Length 537;  
Best Local Similarity 27.5%; Pred. No. 0.89;  
Matches 38; Conservative 6; Mismatches 61; Indels 33; Gaps 5;  
OY 9 PTMTILQAPTTPAPSTIPGRSGPEITFDPLPEPAAAPAGRPSASRGHRKSRRLVLP 68  
DB 38 PTFKECHAPTPEEHAPPKKEHAP-----PKKEHAPAPAAETPPAPCHPDAPQAPAP 92  
OY 69 RVYRQLPVEEPNPAKRLLLLTIVFCQILMAEGVPAPLPDPEDAPNAASLAPTPVSPV 128  
DB 93 AAHEAPTPTHEAAPAH-----EEG-PPPAAPAEAP-----APEP----- 125  
OY 129 LEFFNLTSFSDYALDLS 146  
DB 126 -EPEKPEEPSVPLSLA 142  
-----  
RESULT 13  
XP2\_XENLA  
ID XP2\_XENLA STANDARD: PRT: 439 AA.  
AC P17437; Q08944;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE SKIN SECRETORY PROTEIN XP2 PRECURSOR (APEP PROTEIN).  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;  
OC Xenopodinae; Xenopus.  
RN [1]  
RP SEQUENCE OF 1-25 AND 344-439 FROM N.A.  
RX MEDLINE=92332564; PubMed=1629230;  
RA Hauser F., Roeben C., Hoffmann W.

RT "xp2, a new member of the P-domain peptide family of potential growth  
RT factors, is synthesized in Xenopus laevis skin."  
RL J. Biol. Chem. 267:14451-14455(1992).  
RN [2]  
RP SEQUENCE OF 3-439 FROM N.A.  
RC TISSUE=SKIN;  
RX MEDLINE=90127399; PubMed=2298293;  
RA Gmachl M., Berger H., Thalhammer J., Kreil G.;  
RT "dermal glands of Xenopus laevis contain a polypeptide with a highly  
RT repetitive amino acid sequence."  
RL FEBS Lett. 260:145-148(1990).  
CC -I- FUNCTION: MAY ACT AS A GROWTH FACTOR IN THE GERMINAL LAYER OF THE  
CC EPIDERMIS. MAY ALSO BE INVOLVED IN GROWTH OF REGENERATING GLANDS  
CC AND IN PROTECTION OF THE SKIN FROM THE EXTERNAL ENVIRONMENT.  
CC -I- SUBCELLULAR LOCATION: SECRETED.  
CC -I- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS OF THE PROTEINS ARE  
CC PRODUCED BY ALTERNATIVE SPLICING. THE APEP FORM INCLUDES A LARGE  
CC DOMAIN WITH 33 SEQUENCE REPEATS.  
CC -I- TISSUE SPECIFICITY: SKIN.  
CC -I- SIMILARITY: CONTAINS 2 P-TYPE (TREFOL) DOMAINS.  
CC -I- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 392  
CC ONWARD AND IS SHORTER (418 AA) DUE TO A FRAMESHIFT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M90095; AAA50001.1; .  
DR EMBL; X51394; CAA35759.1; ALT\_FRAME.  
DR PIR; S07498; SKXLG.  
DR PIR; A37331; A37331.  
DR HSP; P04155; 1PS2.  
DR INTERPRO: IPR000319; .  
DR PFAM; PF00088; trefol; 2.  
DR PRINTS; PR00680; PTREFOL.  
DR PROSITE; PS00025; P\_TREFOL; 2.  
KW Signal; Growth factor; Alternative splicing; Amphibian skin; Repeat.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 439 SKIN SECRETORY PROTEIN XP2.  
FT MOD\_RES 23 23 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).  
FT DOMAIN 26 343 33 X REPEATS OF G-[GE]-[AP](2,4)-A-E.  
FT DOMAIN 350 391 P-TYPE 1.  
FT DOMAIN 397 438 P-TYPE 2.  
FT DISULFID 351 377 BY SIMILARITY.  
FT DISULFID 361 376 BY SIMILARITY.  
FT DISULFID 371 388 BY SIMILARITY.  
FT DISULFID 398 424 BY SIMILARITY.  
FT DISULFID 408 423 BY SIMILARITY.  
FT DISULFID 418 435 BY SIMILARITY.  
FT VARSPIC 26 343 MISSING (IN ISOFORM XP2).  
FT CONFLICT 3 3 H -> S (IN REF. 2).  
FT CONFLICT 18 18 C -> W (IN REF. 2).  
SQ SEQUENCE 439 AA; 41173 MW; 38C4A4B57CBAE778 CRC64;  
-----  
Query Match 11.5%; Score 95; DB 1; Length 439;  
Best Local Similarity 25.7%; Pred. No. 0.8;  
Matches 35; Conservative 6; Mismatches 57; Indels 38; Gaps 5;  
OY 15 QAPTAP-----STIPGRSGPEITFDPLPEPAAAPAGRPSASRGHRKSRRLVLP 70  
DB 247 EAPAPAPAEAPAPAPAEAPAEAPAEAPAEAPAEAPAEAPAEAPAEAPAEAPAEAP 301  
OY 71 VRRQLPVE-----EPNPAKRLLLLTIVFCQILMAEGVPAPLPDPEDAPNAASLAPTP 124  
DB 302 AEAPAPAPAEAGGAPSP-----AEGGAPAAAEAGGAPAPAPAE 343  
OY 125 VSPVLE-----PFNLT 135

```
FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
SQ SEQUENCE 620 AA: 65406 MW: 641DD2278AB28524 CRC64;

Query Match 12.2%; Score 100; DB 1; Length 620;
Best Local Similarity 24.3%; Pred. No. 0.46; 75; Indels 44; Gaps 6;
Matches 42; Conservative 12; Mismatches 12;

QY 4 SRSCHTMTILQAPTAPSTIPGPRRGSGEIFTFDPLPEPAAAPAGPSASRGHRKRSR 63
DB 173 SRHGHPPTTYAQPPTPIYSPQVQPPP---TYSPPPTHVQTPSP-PSRGHQPPQ 228
QY 64 RVLY-PRVVRROLVPEENPAKRLFLLLTIVFCOILMAEGVPAP-----LPED 113
DB 229 TRHAPPTHRRAPPHQPSPLHL-----PPSRRQOPPTYSPPPP 270
QY 114 A-----PNAASLAPTPVSPVLEPFLNLTSEPSDYALDLSTELQHPA 154
DB 271 AYAQSPQSPPTYSPPPTYSPPPPPIYSPPPPAYSPPTPTTFSPPPPA 323

RESULT 10
Y091_NPVOP STANDARD; PRT; 279 AA.
AC O10341;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE HYPOTHETICAL 29.3 KDA PROTEIN (ORF92).
OS Orygia pseudotsugata multicapsid polyhedrosis virus (OpMPNV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
NC Nucleopolyhedrovirus.
CC [1]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14112; CAA32311.1; -
KW Repeat.
FT DOMAIN 2911 2980 35 X 2 AA TANDEM REPEATS OF P-Q.
SQ SEQUENCE 3164 AA: 335857 MW: CC5D31FF4F9FE3F4 CRC64;

Query Match 12.0%; Score 98.5; DB 1; Length 3164;
Best Local Similarity 27.3%; Pred. No. 3.1;
Matches 47; Conservative 7; Mismatches 57; Indels 61; Gaps 8;

QY 2 CHRSCHTMTILQA--PTPAPSTIPGP-----RRSGGEIFTFDPLPEPAAAP----- 48
DB 2612 CAPRPPGPAVFAAREARCPGPAESTRAPVGPDRDRLRLSPQ---SSPAPDATAPRPPA 2668
QY 49 -----AGRPSASRGHRKRSRVL-----YPRVV-----RRQLPVEE 79
DB 2669 SSRASASSSSGSRARRRRARRSLARATOAATTOGWRPPALPDVAPVTDFAAPPAPKP 2728
QY 80 PNPARKRLFLLLTIVFCOILMAEGVPAPLPPEPAPNAASLAPTPVSPVLEP 131
DB 2729 PEPAPHALV-----SGVPLPLGPPQAAGQAS--PALPIDPVPPP 2764

RESULT 12
MYPH_CHICK STANDARD; PRT; 537 AA.
AC Q05623;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MYOSIN-BINDING PROTEIN H (MYBP-H) (H-PROTEIN) (86 KDA PROTEIN).
GN MYBP-H.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
CC [1]
CC SEQUENCE FROM N.A., AND SEQUENCE OF 2-37.
RP TISSUE-PECTORALIS MUSCLE;
RX MEDLINE-93155224; PubMed-7679114;
RA Vaughan K.T., Weber F.E., Einheber S., Fischman D.A.;
```



FT ACT\_SITE 181 181 BY SIMILARITY.  
FT MOD\_RES 218 218 PHOSPHORYLATION (ACTIVATES THE KINASE)  
FT (BY SIMILARITY).  
FT MOD\_RES 220 220 PHOSPHORYLATION (ACTIVATES THE KINASE)  
FT (BY SIMILARITY).  
FT DOMAIN 337 340 POLY-ALA.  
FT DOMAIN 433 464 PRO-RICH 1.  
FT DOMAIN 520 523 POLY-ARG.  
FT DOMAIN 577 699 PRO-RICH 2.  
FT CONFLICT 19 32 AREGTRPHRLCS -> GPKVPEAHTAASVA (IN  
REF. 2).  
FT CONFLICT 609 609 L -> V (IN REF. 2).  
FT SEQUENCE 815 AA; 88636 MW; 379AD69803207CCF CRC64;  
Query Match 12.6%; Score 104; DB 1; Length 815;  
Best Local Similarity 24.2%; Pred. No. 0.29;  
Matches 48; Conservative 23; Mismatches 57; Indels 70; Gaps 10;  
QY 4 SRSCPTMTILOQTPAPSTIPGP-----RRSGGPEI-----FTFDPLPEPAAP 48  
Db 575 ARPAAPALTSVPAPAPAPTPTVPQTSPPPGPLAQTPGQPSAGSTSGVPQPACPP 634  
QY 49 AG---RPSASRGHRKRSRVLYPRVVRQLPVEEPNPAKRLLLLTIVFCOILMAEGV 105  
Db 635 PGAPHTPGPG-----PI--PVPAPPQIATSTSLAAGSLVPPPPGL 674  
QY 106 PAP-----LPPEDAPNA--ASLAPTP-----VSPVLEP-FNLTSE 137  
Db 675 PGSSPGVLFPFPGPLPPDAGCAPQSSSESDVNLVTOOLSKSQVEDPLPVPFSGPK 734  
QY 138 PS-----DYALDSTFLQ 151  
Db 735 GSGAGYGVGFDEELFNLQ 752  
RESULT 6  
MODD\_MYCAV STANDARD; PRT; 381 AA.  
AC Q48919;  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE MOLYBDATE UPTAKE SECRETED PROTEIN PRECURSOR (FIBRONECTIN ATTACHMENT  
DE PROTEIN) (FAP-A).  
GN MODD.  
OS Mycobacterium avium.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=101;  
RX MEDLINE=97011577; PubMed=8858587;  
RA Schorey J.S., Holsti M.A., Ratliff T.L., Allen P.M., Brown E.J.;  
RT "Characterization of the fibronectin-attachment protein of  
RT Mycobacterium avium reveals a fibronectin-binding motif conserved  
RT among mycobacteria";  
RL Mol. Microbiol. 21:321-329(1996).  
CC !- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM  
CC MODABC FOR MOLYBDENUM (BY SIMILARITY).  
CC !- SUBCELLULAR LOCATION: SECRETED.  
CC !- SIMILARITY: BELONGS TO THE MODD FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U53585; AAB50543.1; .  
CC Molybdenum; Transport; Signal.

FT SIGNAL 1 32 POTENTIAL.  
FT CHAIN 33 381 MOLYBDATE UPTAKE SECRETED PROTEIN.  
SQ SEQUENCE 381 AA; 38210 MW; 0DACB9A22AA11D2B CRC64;  
Query Match 12.5%; Score 102.5; DB 1; Length 381;  
Best Local Similarity 30.2%; Pred. No. 0.18;  
Matches 42; Conservative 8; Mismatches 42; Indels 47; Gaps 8;  
QY 15 QAPTAP---STIPGRGSGPEIFTFDPLPEPA--AAPAGRPS--ASRGHRKRSRVLYPR 69  
Db 42 EVPTVPSTATTTPHRRRRIP-----GQADNAAGAPAPAPNCGQORPRR----- 88  
QY 70 VVRQLPVEEPNPAKRLLLLTIVFCOILMAEGVPALEPEDAPNAASLAPTPVSPVL 129  
Db 89 --RRMIPTRAPPPA-----GAPPNGAPPAAPNGA--PPPPVDPNA 124  
QY 130 EP-----FNLTSEPSDYAL 143  
Db 125 PPPPADPNAGRIPNSYVL 143  
RESULT 7  
HRX\_MOUSE STANDARD; PRT; 3866 AA.  
ID HRX\_MOUSE  
AC P55200;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ZINC FINGER PROTEIN HRX (ALL-1) (FRAGMENT).  
GN MLL OR HRX OR ALL1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX STRAIN=C57BL/6J, AND B6/CBA; TISSUE=SPLEEN, AND LUNG;  
RC MEDLINE=93317679; PubMed=8327517;  
RA Ma Q., Alder H., Nelson K.K., Chatterjee D., Gu Y., Nakamura T.,  
RA Canaan E., Croce C.M., Siracusa L.D., Buchberg A.M.;  
RT "Analysis of the murine All-1 gene reveals conserved domains with  
RT human ALL-1 and identifies a motif shared with DNA  
RT methyltransferases";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:6350-6354(1993).  
CC !- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.  
CC MAY REGULATE GENES INVOLVED IN SKELETAL FORMATION DURING  
CC EMBRYONIC DEVELOPMENT.  
CC !- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC !- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC !- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION  
CC FACTORS.  
CC !- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
CC !- SIMILARITY: CONTAINS 1 'SET' DOMAIN.  
CC !- SIMILARITY: CONTAINS 3 PHD-FINGER DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L17069; AAA62593.1; .  
CC MGD; MGI:96995; MLL.  
DR INTERPRO: IPR001214; .  
DR INTERPRO: IPR001487; .  
DR INTERPRO: IPR001965; .  
DR INTERPRO: IPR002857; .  
DR PFAM: PF00628; PHD; 3.  
DR PFAM: PF00856; SET; 1.  
DR PFAM: PF02008; ZF-CXXC; 1.

CC BY-NC-ND 4.0 International license.

1. **INTRODUCTION**

F<sup>1</sup>T BINDING

FT BINDING 83 83 ATP (BY SIMILARITY).

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL: S81914; AAB36278.1; -;  
CC DR EMBL: X96438; CAA65304.1; -;  
CC DR EMBL: Y14551; CAA74886.1; -;  
CC DR EMBL: T49531; -; NOT\_ANNOTATED\_CDS.  
CC DR MIM: 602996; -;  
CC KW Glycoprotein; Transmembrane; Signal-anchor.  
CC FT DOMAIN 1 82 CYTOPLASMIC (POTENTIAL).  
CC FT TRANSMEM 83 99 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
CC FT (POTENTIAL).  
CC FT DOMAIN 100 156 \* EXTRACELLULAR (POTENTIAL).  
CC FT CARBOHYD 133 133 N-LINKED (GLCNAC).  
CC FT CONFLICT 54 54 A -> G (IN REF. 1).  
CC FT CONFLICT 106 106 P -> R (IN REF. 1).  
CC SQ SEQUENCE 156 AA; 16928 MW; 83C06116C80B9240 CRC64;

Query Match 100.0%; Score 823; DB 1; Length 156;  
Best Local Similarity 100.0%; Pred. No. 1.6e-58;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTIQAAPTAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASGRHK 60  
DB 1 MCHSRCHPTMTIQAAPTAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASGRHK 60  
QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120  
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120

QY 121 APTVPVLEPFLNTSEPSDYALDLSLTFLOQHPAAF 156  
DB 121 APTVPVLEPFLNTSEPSDYALDLSLTFLOQHPAAF 156

RESULT 2  
ID IEX1\_MOUSE STANDARD; PRT; 153 AA.  
AC P46694;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY  
DE PROTEIN GLY96) (IMMEDIATE EARLY RESPONSE 3 PROTEIN).  
GN IER3 OR IEX1 OR GLY96.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C;  
RX MEDLINE=93173526; PubMed=8437864;  
RA Charles C.H., Yoon J.K., Sinks J.S., Lau L.F.;  
RT "Genomic structure, cDNA sequence, and expression of gly96, a growth  
RT factor-inducible immediate-early gene encoding a short-lived  
RT glycosylated protein.";  
RL Oncogene 8:797-801(1993).  
CC -!- FUNCTION: NOT KNOWN; EXPRESSED DURING THE G0-G1 TRANSITION OF THE  
CC CELL CYCLE.  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).  
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE LUNG, TESTES  
CC AND THE UTERUS.  
CC -!- INDUCTION: BY SERUM GROWTH FACTORS AND TPA.  
CC -!- PTM: GLYCOSYLATED.  
CC -!- SIMILARITY: STRONG, TO HUMAN ORTHOLOG.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL: X67644; -; NOT\_ANNOTATED\_CDS.  
CC DR PIR: S33363; S33363.  
CC DR MGD: MGI:104814; IER3.  
CC KW Glycoprotein; Transmembrane; Signal-anchor.  
CC FT DOMAIN 1 85 CYTOPLASMIC (POTENTIAL).  
CC FT TRANSMEM 86 102 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
CC FT (POTENTIAL).  
CC FT DOMAIN 103 153 EXTRACELLULAR (POTENTIAL).  
CC FT CARBOHYD 137 137 N-LINKED (GLCNAC).  
CC SQ SEQUENCE 153 AA; 16875 MW; F4429E3B0120DE11 CRC64;

Query Match 63.8%; Score 525; DB 1; Length 153;  
Best Local Similarity 70.3%; Pred. No. 4.2e-35;  
Matches 109; Conservative 9; Mismatches 29; Indels 8; Gaps 2;

QY 1 MCHSRCHPTMTIQAAPTAPSTIPGPRGSGPEIFTFDPLPEPAAAPAGRPSASGRHK 60  
DB 1 MCHSRCHPTMTIQAAPTAPSTIPGPRGSGPEIFTFDPLPERAVVSTARLNTSRGRHK 60  
QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPDPEDAPNAASL 120  
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVAPLPEDATSA--V 118  
QY 121 APTVPS-----PVLEPFLNTSEPSDYALDLSLTF 149  
DB 119 TPEISAPITAPPVLEPLNTSEPSDYALDLSLTF 153

RESULT 3  
ID EXTN\_SORBI STANDARD; PRT; 283 AA.  
AC P24152;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE EXTENSIN PRECURSOR (PROLINE-RICH GLYCOPROTEIN).  
GN HRGP.  
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LEAF;  
RX MEDLINE=91370882; PubMed=1893107;  
RA Raz R., Gretin C., Puigdomenech P., Martinez-Izquierdo J.A.;  
RT "The sequence of a hydroxyproline-rich glycoprotein gene from Sorghum  
RT vulgare.";  
RL Plant Mol. Biol. 16:365-367(1991).  
CC -!- FUNCTION: STRUCTURAL COMPONENT IN PRIMARY CELL WALL.  
CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE  
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN  
CC GLYCOSYLATED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL: X56010; CAA39485.1; -;  
CC DR PIR: S14449; S14449.  
CC DR INTERPRO: IPR002965;  
CC DR PRINTS: PR01217; PRICEXTENS.  
CC KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;  
CC Hydroxylation.

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 22:05:39 ; Search time 43.61 Seconds  
(without alignments)  
115.521 Million cell updates/sec

Title: US-08-799-910-10  
Perfect score: 823  
Sequence: 1 MCHRSRCHPTTILQAPTPA.....EPSDYALDLSTFLQHPAAAF 156

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	823	100.0	156	1	IEXL_HUMAN
2	525	63.8	153	1	IEXL_MOUSE
3	112.5	13.7	283	1	EXTN_SORBI
4	105	12.8	3149	1	TEGU_EBV
5	104	12.6	815	1	MR07_HUMAN
6	102.5	12.5	381	1	MODD_MYCAV
7	102.5	12.5	3866	1	HRX_MOUSE
8	101.5	12.3	268	1	NO20_MEDTR
9	100	12.2	620	1	EXTN_TOBAC
10	99.5	12.1	279	1	Y091_NPVOP
11	98.5	12.0	3164	1	TEGU_HSV11
12	95.5	11.6	537	1	MYPH_CHICK
13	95	11.5	439	1	XP2_XENLA
14	94.5	11.5	510	1	HRN_HUMAN
15	94.5	11.5	758	1	HUNB_DROME
16	94.5	11.5	1595	1	SOS_DROME
17	94	11.4	1290	1	PER1_HUMAN
18	93.5	11.4	1402	1	IF4G_RABIT
19	93	11.3	817	1	VRPL_YEAST
20	92.5	11.2	605	1	BRL1_EBV
21	92.5	11.2	625	1	NIFA_AZOBR
22	92	11.2	534	1	APG_ARATH
23	92	11.2	3969	1	HRX_HUMAN
24	91.5	11.1	267	1	EXTN_MAIZE
25	91.5	11.1	615	1	MUTL_ECOLI
26	91.5	11.1	1133	1	SREL_CRIGR
27	90	10.9	316	1	CDNC_HUMAN
28	90	10.9	449	1	APG_BRANA
29	89.5	10.9	1487	1	ICP4_HSVK
30	89.5	10.9	2715	1	TRX2_HUMAN
31	89	10.8	1291	1	PER1_MOUSE
32	88.5	10.8	865	1	CPN_DROME
33	88	10.7	474	1	VTP3_TTV1V

34	88	10.7	477	1	BLAR_HUMAN	P08588	homo sapien
35	87.5	10.6	543	1	VP61_NPVAC	Q03209	autographa
36	87.5	10.6	591	1	MNT_MOUSE	O08789	mus musculus
37	87	10.6	1575	1	SYJ1_HUMAN	O43426	homo sapien
38	87	10.6	2414	1	P300_HUMAN	O09472	homo sapien
39	86.5	10.5	1023	1	SREL_RAT	P56720	rattus norv
40	86.5	10.5	1487	1	ICP4_HSVB	P28925	equine herp
41	86	10.4	390	1	VG14_HSV11	P06487	herpes simp
42	86	10.4	1324	1	IRS2_HUMAN	Q9Y4H2	homo sapien
43	85.5	10.4	224	1	Y091_NPVAC	P41479	autographa
44	85.5	10.4	626	1	GPBA_HUMAN	P07359	homo sapien
45	85.5	10.4	1233	1	NME3_HUMAN	Q14957	homo sapien

## ALIGNMENTS

RESULT 1	
IEX1_HUMAN	
ID	IEX1_HUMAN STANDARD; PRT; 156 AA.
AC	P46695; Q93044; Q92691;
DT	01-NOV-1995 (Rel. 32, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)
DT	30-MAY-2000 (Rel. 39, Last annotation update)
DE	RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY
DE	PROTEIN GLY96) (IMMEDIATE EARLY RESPONSE 3 PROTEIN) (PACAP-RESPONSIVE
DE	GENE 1 PROTEIN) (PRGI PROTEIN) (DIFFERENTIATION-DEPENDENT GENE 2
DE	PROTEIN) (DIF-2 PROTEIN).
GN	IEX3 OR IEX1 OR PRG1 OR DIF2.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=PLACENTA;
RX	MEDLINE=96181295; PubMed=8603392;
RA	Kondratyev A.D., Chung K.-N., Jung M.O.;
RT	"Identification and characterization of a radiation-inducible
RT	glycosylated human early-response gene.";
RL	Cancer Res. 56:1498-1502(1996).
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=96221139; PubMed=8653710;
RA	Schaefer H., Trauzold A., Siegel E.G., Folsch U.R., Schmidt W.E.;
RT	"PRG1: a novel early-response gene transcriptionally induced by
RT	pituitary adenylate cyclase activating polypeptide in a pancreatic
RT	carcinoma cell line.";
RL	Cancer Res. 56:2641-2641(1996).
RN	[3]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=9739426; PubMed=9196025;
RA	Pietzsch A., Buechler C., Aslanidis C., Schmitz G.;
RT	"Identification and characterization of a novel monocyte/macrophage
RT	differentiation-dependent gene that is responsive to
RT	lipopolysaccharide, ceramide, and lysophosphatidylcholine.";
RL	Biochem. Biophys. Res. Commun. 235:4-9(1997).
RN	[4]
RP	PRELIMINARY SEQUENCE OF 1-106 FROM N.A.
RC	TISSUE=PLACENTA;
RA	Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman M.,
RA	Hultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J.,
RA	Rifkin L., Rohlfing T., Tan F., Trevaskis E., Waterston R.,
RA	Williamson A., Wohldmann P., Wilson R.;
CC	Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC	-1- INDUCTION: BY RADIATION, TPA, OKADAIC ACID AND TNF-ALPHA.
CC	-1- PTM: GLYCOSYLATED.
CC	-1- SIMILARITY: STRONG, TO MOUSE ORTHOLOG.
CC	-1- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC	FRAMESHIFTS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation.

---



A; Reference number: A56708; MUID: 95279403

A;Status: preliminary

A: Residues: 1-815 <2HC

A; Cross-references: GR:

**C;Keywords:** ATP

F;60-68/Region: protein

Best: Local similarity

[illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

QY 49 AG--RPSASRGH

Db 635 PCPAPHPTGPPG-

QY 106 PAP-----

Db 675 PGSSTPGVLPYFP

07 138 DS-----DVATDIEN

—

## RESULT 15

9T727216  
hypothetical protein

C; Species: *Caenorhabditis*

C:Accession: T22216

submitted to the EMBL Data

A: Accession. T22216

A; Status: preliminary; t

A;Residues: 1-241 <WIL>

A; Experimental source: c

A:Gene: CESP:F45B8 3

A; Map position: X  
A; Tetrone: 31/3. 69/3.

Query Match

Matches 41: Conserv

Q. Now, you're going to ask me to read this, is that right?

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841

59 KKSKRVLYPRVVR

131 -----PI,VCC

119 SLAPTPVSPVI,EP

164 VAApVPTNPCCOP

A:Residues: 1-351 <WOE>  
A:Cross-references: EMBL:L29028; NID:g530877; PIDN:AAB53954.1; PID:g530878

Query Match 12.8%; Score 105; DB 2; Length 351;  
Best Local Similarity 24.2%; Pred. No. 0.19;  
Matches 30; Conservative 16; Mismatches 56; Indels 22; Gaps 3;  
  
QY 9 PTMTILOATTPAPSTIPGRRGSGPEIFTFDLPPEAAAPAGRPASRGHRRSRRLVLP 68  
Db 172 PPSPPSPSPSPSPSPKASPSPP-KASPSPPKASPSPPKASPSPPSPSPSPSPSP 230  
QY 69 RVVRRQLPVEEPNPAKRLLFLLLTIVFCOILMAEEGVPAIP-PEDAPNAASLAPTPVSP 127  
Db 231 KASPNASPOOSTPSPR-----PSPTSPSPSPKASPPSPSPSPSPSP 270  
QY 128 VLEP 131  
Db 271 SASP 274

RESULT 11  
T17525  
proline-rich protein A35L - Chlorella virus PBCV-1  
C:Species: Chlorella virus PBCV-1  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T17525  
R:Graves, M.V.; Van Etten, J.L.  
submitted to the EMBL Data Library, May 1999  
A:Reference number: Z18806  
A:Accession: T17525  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-549 <GRA>  
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96403.1  
A:Experimental source: specific host Chlorella strain NC64A  
C:Genetics:  
A:Gene: A35L

Query Match 12.8%; Score 105; DB 2; Length 549;  
Best Local Similarity 27.8%; Pred. No. 0.3;  
Matches 35; Conservative 11; Mismatches 36; Indels 44; Gaps 6;  
  
QY 16 APTAPSTIPGRRGSGPEIFTFDLPPEAAAPAGRPASRGHRRSRRLVPRVRRQL 75  
Db 445 APKPAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 485  
QY 76 PVVEPNPAKRLLFLLLTIVFCOILMAEEGVPAIP-PEDAPNAASLAPTPV-SPVLEPN 133  
Db 486 PKPAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 522  
QY 134 LTSEPS 139  
Db 523 ITSQVS 528

RESULT 12  
QORE8  
hPLF1 protein - human herpesvirus 4 (strain B95-8).  
C:Species: human herpesvirus 4, Epstein-Barr virus  
C:Date: 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change 16-Jul-1999  
C:Accession: G93065; A03747; S32993  
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.  
Mol. Biol. Med. 1, 21-45, 1983  
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus  
A:Reference number: A93065; MUID:85035713  
A:Accession: G93065  
A:Molecule type: DNA  
A:Residues: 1-3149 <BAN>  
A:Cross-references: EMBL:V01555; NID:g59074; PIDN:CAA24839.1; PID:g1334853  
R:Bankier, A.T.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H  
Nature 310, 207-211, 1984

A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.  
A:Reference number: A03794; MUID:84270667  
A:Contents: annotation; protein coding region  
C:Superfamily: human herpesvirus 4 BPLF1 protein

Query Match 12.8%; Score 105; DB 1; Length 3149;  
Best Local Similarity 25.3%; Pred. No. 1.7;  
Matches 44; Conservative 21; Mismatches 51; Indels 58; Gaps 9;  
  
QY 9 PTMTILOATTPAPS-----TIPGRRG-SGPEIFTFDLPPEAAAPAGRPASRGHRRKS 62  
Db 484 PVSTIAPSVTPSPRLPIQIPLPQAAPNSNKPIPTTSPSTAAAA--PTTT----- 534  
QY 63 RRVLYPRVRRQ-----LPVEEPNPAKRLLFLLLTIVFCOILMAEEGVPAIP-- 108  
Db 535 --TLPSPPTTQOOPQOSAAAPAPSPLLPQQQPTPS-----ANAPSP 572  
QY 109 -LPPEADAPNAASLAPTPVSPVLEPNLTS---EPSDYALDLSTFIO---QHFAA 155  
Db 573 LLPOQQPPPSAARAPSPLPQQOQLPSATPAPPAQQLPPSATTLEPEKNIPPA 626

## RESULT 13

S76953  
protein kinase (EC 2.7.1.-), 58K - Synecocystis sp. (strain PCC 6803)  
N:Alternate names: protein slr0599  
C:Species: Synecocystis sp.  
A:Variety: PCC 6803  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: S76953  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S76953  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-535 <KAN>  
A:Cross-references: EMBL:D90917; GB:AB001339; NID:g1653836; PIDN:BAAL8865.1; PLOC:  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Superfamily: Synecocystis sp. protein kinase, 58K; protein kinase homology  
C:Keywords: phosphotransferase; protein kinase  
E:10-274/Domain: protein kinase homology <KIN>

Query Match 12.7%; Score 104.5; DB 1; Length 535;  
Best Local Similarity 25.6%; Pred. No. 0.32; Indels 41; Gaps 4;  
Matches 32; Conservative 9; Mismatches 43;  
  
QY 16 APTAPSTIPGRRGSGPEIFTFDP-----LPPEAAAPAGRPASRGHRRSRRLVYPR 69  
Db 411 SPEPTSP 456  
QY 70 VVRQLPVEEPNPAKRLLFLLLTIVFCOILMAEEGVPAIP-PEDAPNAASLAPTPVSPV 128  
Db 457 -----PTISPPSPSTISIPV-----TPAPVPKPSPTPKPTVPQISPT 496  
QY 129 LEFPN 133  
Db 497 PQPSN 501

## RESULT 14

B56708  
extracellular signal-regulated kinase 5 - human  
C:Species: Homo sapiens (man)  
C:Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 24-Sep-1999  
C:Accession: B56708  
R:Zhou, G.; Bao, Z.Q.; Dixon, J.E.  
J. Biol. Chem. 270, 12665-12669, 1995

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-496 <GRA>  
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96773.1  
A:Experimental source: specific host *Chlorella* strain NC64A  
C:Genetics:  
A:Note: A405R

Query Match 13.0%; Score 107; DB 2; Length 496;  
Best Local Similarity 22.8%; Pred. No. 0.19;  
Matches 37; Conservative 20; Mismatches 57; Indels 48; Gaps 6;  
QY 16 APTAPSTIPGRRSGPEIFTDPLPEAAAPAGRSASRGHKKSRRLVYPRVRRQL 75  
DB 128 APKAPAPAPKAPKSPK\*---PAKPAKAPKAPKAPKAPKAPKAPKAPKAPK 180  
QY 76 PVEEPNPKRLFLLLTVFCQILMAEGVPAPLP-----PEDAPNAASLAP----- 122  
DB 181 PKPAKPAK\*-----PAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 226  
QY 123 -----TPVSPVLE-----PFLTSEPSDYALDLSTFLQ 150  
DB 227 DINDKAPMLPIEDIPSSFQDQTOFTVTSITTTATNIDLPGYVK 268

RESULT 7  
A49227  
sialidase - Actinomyces viscosus  
C:Species: Actinomyces viscosus  
C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 22-Oct-1999  
C:Accession: A49227  
R:Yeung, M.K.  
Infect. Immun. 61, 109-116, 1993  
A:Title: Complete nucleotide sequence of the Actinomyces viscosus T14V sialidase gene: P  
A:Reference number: A49227; MUID:93114861  
A:Contents: T14V  
A:Accession: A49227  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-901 <YEU>  
A:Cross-references: GB:L06898; NID:g289087; PIDN:AAA21932.1; PID:g141852  
A:Note: sequence extracted from NCBI backbone (NCBI:121598, NCBI:P:121599)

Query Match 12.9%; Score 106.5; DB 2; Length 901;  
Best Local Similarity 29.0%; Pred. No. 0.37;  
Matches 45; Conservative 19; Mismatches 64; Indels 27; Gaps 8;  
QY 9 PTWTILOAPYPAVSTIPGRRSGPEIFTDPLPEAAAPAGRSASRGHKKSRRLVYP 68  
DB 722 PTTAPSTERTAPAPSSAPEQTDGP---TAAPAPETSSAPAAEPTQA-----PTVAP 770  
QY 69 RVVRRQIPVEEPNPKR--LFLFLLLTVFCQILMA--EGVP-----APLPEDAPNAASL 120  
DB 771 SVPTGAPGAPSSAPKPGATGRAPSVVNPATGAATEPTGTPSSASAPSRNA----- 824  
QY 121 APTPVSPVLEPFLNLTSEPSDYALDLSTFLQHPAA 155  
DB 825 APTP-KPGMEPDEI-DRPSDGTMAOPTGGASAPSA 857

RESULT 8  
T00037  
hypothetical protein KIAA0284 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 05-Nov-1999  
C:Accession: T00037  
R:Ohara, O.; Nagase, T.; Ishikawa, K.; Nakajima, D.; Ohira, M.; Seki, N.; Nomura, N.  
submitted to the EMBL Data Library, August 1997  
A:Description: Prediction of the coding sequences of unidentified human genes.  
A:Reference number: 214073  
A:Accession: T00037

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1323 <OHA>  
A:Cross-references: EMBL:AB006622; NID:d1170675; PIDN:BAA22953.1; PID:d1C23829  
A:Experimental source: brain  
C:Genetics:  
A:Note: KIAA0284

Query Match 12.9%; Score 106.5; DB 2; Length 1323;  
Best Local Similarity 26.8%; Pred. No. 0.54;  
Matches 45; Conservative 13; Mismatches 67; Indels 43; Gaps 7;  
QY 3 HSRSCHTMTILOAPTPA-----PSTIPGRRSGPEIF-----TFDPL--PEPAA 46  
DB 368 HSPSGDPKADKRRGTPADRDPSVPAPVAGSGSPORAGSLKREKTEERLGSPPAS 427  
QY 47 APAGRSASRGHKKSRRLVYPRVRRQIPVEEPNPKRLFLLLTVFCQILMAEGVP 106  
DB 428 RTPARFGSVGRKSLAQDFMAOCLRESSPAARPSPEK-----VP 467  
QY 107 APLPPEDAPNAASLA--PTPVSPVLEPFLNLT-----SEPSDYALDLSTF 148  
DB 468 PVLPAPLTHGTSVPGPTPPAPTDP-QLTKARKQEDDLSLDAGTY 514

RESULT 9  
T30351  
mucin-like protein - Lymantria dispar nuclear polyhedrosis virus  
C:Species: Lymantria dispar nuclear polyhedrosis virus, LdMNPV  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T30351  
R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.;  
Virology 253, 17-34, 1999  
A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lym  
A:Reference number: 220836; MUID:99124785  
A:Accession: T30351  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1029 <KUZ>  
A:Cross-references: EMBL:AF081810; PIDN:AAC70189.1

Query Match 12.9%; Score 106; DB 2; Length 1029;  
Best Local Similarity 29.5%; Pred. No. 0.46;  
Matches 51; Conservative 6; Mismatches 52; Indels 64; Gaps 8;  
QY 16 APTAPSTIPGRRSGGP-EIETFF-----DPLPEAAAPAGRPSASRCHKRSRRLV 67  
DB 637 APEAPEPVPEP---SAPVDYFTLSAEPAPEPAPEPAPEPAPEPAPEPAPEP 593  
QY 68 PRVRRQIPVEEPNPKRLFLLLTVFCQILMAEGVPAPLPEDAPNAASLAPTPVSP 127  
DB 694 P-VEPALEPVE-----PALEPVEPALGPGPALEPGE 725  
QY 128 VLEPNLSE-----PSDYALDL-----STF-----LQHPAA 155  
DB 726 ALEPGESTAEAAASELAVERPAQAPDLTSDSAVESTFGHKSTVASELARPAA 778

RESULT 10  
S50754  
hypothetical protein WP6 - Chlamydomonas eugametos  
C:Species: Chlamydomonas eugametos  
C:Date: 14-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 21-Jul-2000  
C:Accession: S50754  
R:Woessner, J.P.; Molendijk, A.J.; van Egmond, P.; Klis, F.M.; Goodenough, U.W.;  
Plant Mol. Biol. 26, 947-960, 1994  
A:Title: Domain conservation in several volvocalean cell wall proteins.  
A:Reference number: S50754; MUID:95093034  
A:Accession: S50754  
A:Status: preliminary  
A:Molecule type: mRNA



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: January 30, 2001, 21:28:35 ; Search time 21.17 Seconds  
(without alignments)  
500.354 Million cell updates/sec

Title: US-08-799-910-10  
Perfect score: 823  
Sequence: 1 MCHSRSCHPTMTILOAPTPA.....EPSDYALDLSTFLQHPAAF 156  
Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 6790655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	819	99.5	156	2 JC5537	differentiation-de
2	525	63.8	153	2 S33363	gly96 protein - mo
3	112.5	13.7	283	2 S13383	hydroxyproline-ric
4	109.5	13.3	552	2 T08148	proline-rich myros
5	107.5	13.1	581	2 T22341	hypothetical prote
6	107	13.0	496	2 T17908	proline/lysine-ric
7	106.5	12.9	901	2 A49227	sialidase - Actino
8	106.5	12.9	1323	2 T00037	hypothetical prote
9	106	12.9	1029	2 T30351	mucin-like protein
10	105	12.8	351	2 S50754	hypothetical prote
11	105	12.8	549	2 T17525	proline-rich prote
12	105	12.8	3149	1 Q0BE8	BPLFI protein - hu
13	104.5	12.7	535	1 S76953	protein kinase (EC
14	104	12.6	815	2 B56708	extracellular sign
15	103	12.5	241	2 T22216	hypothetical prote
16	102.5	12.5	288	2 T17737	proline-rich prote
17	102.5	12.5	3866	2 B48205	All-1 protein -GTE
18	102.5	12.5	3869	2 A48205	All-1 protein +GTE
19	101	12.3	339	2 T17636	proline-rich prote
20	100	12.2	620	2 S06733	hydroxyproline-ric
21	99.5	12.1	279	2 T10361	hypothetical prote
22	98.5	12.0	3164	1 WMBEH6	UL36 protein - hum
23	98	11.9	473	2 S50755	hypothetical prote
24	97.5	11.8	225	2 T17815	proline-rich prote
25	97.5	11.8	2157	2 S71461	proline-rich prote
26	97.5	11.8	3938	2 T42761	Bassoon protein -
27	97	11.8	461	2 T10741	extensin-like prot
28	96.5	11.7	168	2 S52994	arabinogalactan-li
29	96	11.7	185	2 T00519	proline-rich prote

30	95.5	11.6	524	2 A75588	probable protein k
31	95.5	11.6	537	2 A46611	myosin-binding pro
32	95.5	11.6	753	2 JQ0532	OP protein - Ksne
33	95.5	11.6	1596	2 A41216	guanine nucleotide
34	95	11.5	202	2 JQ0964	hydroxyproline-ric
35	95	11.5	377	2 A48018	mucin 7 precursor,
36	95	11.5	416	1 SKXLAG	dermal gland prote
37	95	11.5	801	2 T29018	hypothetical prote
38	95	11.5	3511	2 A59295	unconventional myo
39	94.5	11.5	510	1 S43692	transcription fact
40	94.5	11.5	758	2 A29253	finger protein hun
41	94.5	11.5	2342	2 T13412	hypothetical prote
42	94	11.4	328	2 JQ0985	hydroxyproline-ric
43	94	11.4	925	2 S27920	nuclear antigen EB
44	94	11.4	1290	2 T00018	period protein hom
45	93.5	11.4	316	2 T31880	hypothetical prote

ALIGNMENTS

RESULT 1  
JC5537  
differentiation-dependent protein DIF-2 - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Sep-1997 #sequence\_revision 05-Sep-1997 #text\_change 07-May-1999  
C:Accession: JC5537  
R:Pietzsch, A.; Buechler, C.; Aslanidis, C.; Schmitz, G.  
Biochem. Biophys. Res. Commun. 235, 4-9, 1997  
A:Title: Identification and characterization of a novel monocyte/macrophage diff  
A:Reference number: JC5537; MUID:97339426  
A:Accession: JC5537  
A:Molecule type: mRNA  
A:Residues: 1-156 <PIF>  
A:Experimental source: monocyte  
A>Note: the authors translated the codon CCG for residue 106 as Arg

Query Match 99.5%; Score 819; DB 2; Length 156;  
Best Local Similarity 99.4%; Pred. No. 1.2e-58;  
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MCHSRCHPTMTILOAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60  
|||||  
DB 1 MCHSRCHPTMTILOAPTPAPSTIPGPRGSGPEITFDPLPEPAAAPAGRPSASRGHRK 60  
QY 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEEGVPAPLPPEDAPNAASL 120  
|||||  
DB 61 RSRRLVPRVRRQLPVEEPNPAKRLFLLLTIVFCOILMAEEGVPAPLPPEDAPNAASL 120  
QY 121 APTPVSPVLEPFLNLTSEPSDYALDLSTFLQHPAAF 156  
|||||  
DB 121 APTPVSPVLEPFLNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 2  
S33363  
gly96 protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Feb-1997  
C:Accession: S33363  
R:Charles, C.H.; Yoon, J.K.; Simske, J.S.; Lau, L.F.  
Oncogene 8, 797-801, 1993  
A:Title: Genomic structure, cDNA sequence, and expression of gly96, a growth fac  
A:Reference number: S33363; MUID:93173526  
A:Accession: S33363  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-153 <CHA>  
A:Cross-references: EMBL:X67644  
C:Genetics:  
A:Introns: 70/3  
C:Keywords: transmembrane protein



```

QY      17 PTPASTIPGPRGGSGPEITFDLPETAAAPAGPSASGRHKRSRRVLYPRVVRRQLP 76
       ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      87 PPPPAKPKPPPKPPAPE-----PEPPPPAPRPAPP-----P 120
QY      77 VEEPNAKRLLFLLLTIVFCQILMAEGVPALPEDAP-----NAASLAPT- ---VSP 127
       ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB     121 APEPSPTP-----TPAPTppvvrpaapapapsvtpttprpslsp 160
QY     128 VLEP 131

```

QY 109 LF --- PEDAFNAASLAPIPV --- SPVLEPFLNTSEPSDYA I42

109 LP-----PEDAPNAASLAPTPV-----SPVLEPFNLTSEPSDYA 142  
| | | |||| | || : | | | | :  
510 APAPAPEAAPAAAPDAAAPAEAPVPVEVAPAPAPAPAPAPAESDASGYS 555

Qy Db

Search completed: January 30, 2001, 23:11:29  
Job time: 3994 sec

```
RX MEDLINE=97127825; PubMed=8972576;
RA Ramon A., Gil R., Bungal M., Sentandreu R., Valentin E.;
RT "A novel cell wall protein specific to the mycelial form of Yarrowia
lipolytica.";
RL Yeast 12:1535-1548(1996).
DR EMBL; 281006; CAB02634.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN 17 285 POTENTIAL.
FT CHAIN 17 285 POTENTIAL.
SQ SEQUENCE 285 AA; 30036 MW; 0BA7FFIC515CD57D CRC64;

Query Match 13.2%; Score 109; DB 3; Length 285;
Best Local Similarity 28.4%; Pred. No. 0.012;
Matches 40; Conservative 11; Mismatches 62; Indels 28; Gaps 6;

OY 1 MCHSRCHPTMTILOAQTAP-STIPGRRGSGPEITFDPLPEPAAAPGRPSASGRHR 59
Db 122 ICHTKVTLSTEVLTPTTPVQTTTAVPEPKPTPEVPEVKPEPTP-EVPGVKPEPTGPP 180
OY 60 KRRRVLYPRVRRQLP-VVEPNPAKRLLFLLLTIVFCQILMAEEGVPAPLPEDAPNAA 118
Db 181 APKPEPEVPEVKPEPTPEVPEVRP-----EPTPAPLPRP--PSLR 218
OY 119 SLAPTPVSPVLEPNLTSEPS 139
Db 219 SLRS---SPSLPLPLPPSPS 236

RESULT 11
O15047 PRELIMINARY; PRT: 1709 AA.
AC O15047;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE KIAA0339 PROTEIN (FRAGMENT).
GN KIAA0339
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
DR EMBL; AB002337; BAA20797.2; -.
DR INTERPRO; IPR000504; -.
DR INTERPRO; IPR01214; -.
DR INTERPRO; IPR002965; -.
DR PFAM; PF000076; rrm; 1.
DR PFAM; PF00856; SET; 1.
DR PRINTS; PR01217; PRICHEXTENS.
FT NON_TER 1
SQ SEQUENCE 1709 AA; 186231 MW; DCF2E0FF716B672C CRC64;

Query Match 13.2%; Score 109; DB 4; Length 1709;
Best Local Similarity 32.0%; Pred. No. 0.069;
Matches 44; Conservative 10; Mismatches 53; Indels 28; Gaps 7;

OY 6 SCUPTMTILOAQTAPSTIPGRRGSGPEITFDPLPEPAAAPGRPSASGRHKRRRV 65
Db 1078 SASPPPEVPVPTAPVEVPVPERVAGSPV---TPLPEQASPA-RPACHTESPPSAPL 1133
OY 66 LYPRVRRRLPVEEPNPAKR-----LLFL-----LTIVFCQILMAEEGVPAPLPP 111
Db 1134 RPP-----EPPAGPAPAPRDERPSSPIPLPPKKRRKTVSFSAI----EVPVPEPP 1184
```

```
OY 112 EDAPNAASLAPTPVS 126
Db 1185 PATPPQAKF-PGPAS 1198

RESULT 12
Q91BT9 PRELIMINARY; PRT: 3325 AA.
AC Q91BT9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE MAJOR TEGUMENT PROTEIN.
GN UL36.
OS Turkey herpesvirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.
OX NCBI_TaxID=10390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GA;
RX MEDLINE=92237304; PubMed=1315048;
RA Jones D., Lee L., Liu J.L., Kung H.J., Tillotson J.K.;
RT "Marek disease virus encodes a basic-leucine zipper gene resembling
the fos/jun oncogenes that is highly expressed in lymphoblastoid
tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4042-4046(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GA;
RA Lee L.F., Wu P., Sui D., Ren D., Kung H.J.;
RT "The Complete UL Sequence of Serotype I Marek's Disease Virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2000).
DR EMBL; AF147806; AAF66771.1; -.
SQ SEQUENCE 3325 AA; 365698 MW; 5D4A84EF719BB9FD CRC64;

Query Match 13.2%; Score 109; DB 12; Length 3325;
Best Local Similarity 25.2%; Pred. No. 0.13;
Matches 31; Conservative 14; Mismatches 52; Indels 26; Gaps 3;

OY 16 APTPAPSTIPGRRGSGPEITFDPLPEPAAAPGRPSASGRHKRRRVLYPRVRRQL 75
Db 2802 SPAPKSPAPKPKPPDPD---FKSPAPKSPASPKSPASPKSPAPKPPAPDSKPS 2858
OY 76 PVEEPNPAKRLLFLLLTIVFCQILMAEEGVPAPLPEDAPNAASLAPTP-----VSPV 128
Db 2859 PAPKPKPPP-----TPDSKPSAPKPKSPASPKPLPVPPPNDSKSTSPV 2902

OY 129 LEP 131
Db 2903 PNP 2905

RESULT 13
Q9L568 PRELIMINARY; PRT: 233 AA.
AC Q9L568;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE PSPA (FRAGMENT).
GN PSPA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=39;
RA Beall B.W.;
RT "psa sequence types from multiresistant pneumococci.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
```



[illegible]



Db 181 DLSTFLQHPAAF 193  
|||||

## RESULT 2

Q9NXH8 PRELIMINARY; PRT: 423 AA.  
AC Q9NXH8;  
DT 01-JAN-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)  
DE CDNA FLJ20245 FIS, CLONE COLF6454.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=COLON MUCOSA;  
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,  
RA Nakamura Y., Isono T., Sugano S.;  
RT "NEDO human cDNA sequencing project."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK000252; BAA91032.1; -  
SQ SEQUENCE 423 AA; 46942 MW; F3F52BC058CAD51E CRC64;

Query Match 14.9%; Score 122.5; DB 4; Length 423;  
Best Local Similarity 42.2%; Pred. No. 0.001; 34; Indels 9; Gaps 5;  
Matches 38; Conservative 9; Mismatches 34; Indels 9; Gaps 5;

QY 19 PAPSTIP-GPRRG-SGPEITFDPLPEAAAPAGRPSASGRHKKRSRRVLYPRVVRQLP 76  
Db 58 PRPGCSAPRADLPQPKFTED---SPAELPSRTP---RKKRRSRLVLYPETSRRKYP 111  
QY 77 -VEEPNPAKRLFLLLTIVFCQILMAEEGV 105  
Db 112 RVEHRSRAQRCLLLVAVIGQVLAENL 141

## RESULT 3

O15026 PRELIMINARY; PRT: 1668 AA.  
AC O15026;  
DT 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)  
DE KIAA0309 (FRAGMENT).  
GN KIAA0309.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=97349984; PubMed=9205841;  
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,  
RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. VII.  
RT The complete sequences of 100 new cDNA clones from brain which can  
RT code for large proteins in vitro."  
RL DNA Res. 4:141-150(1997).  
DR EMBL; AB002307; BAA20768.1; -  
DR INTERPRO; IPR000637; -  
DR INTERPRO; IPR001650; -  
DR PFAM; PF00271; helicase\_C; 1.  
DR PFAM; PF02178; AT\_hook; 3.  
DR PRINTS; PRO0929; ATHOOK.  
FT NON\_TER 1  
SQ SEQUENCE 1668 AA; 175162 MW; 7717C8FC6A396C49 CRC64;

## Query Match

Best Local Similarity 13.9%; Score 114.5; DB 4; Length 1668;  
Matches 41; Conservative 12; Mismatches 12; Indels 37; Gaps 6;  
QY 6 SCHPTMTILQAP--TPAPSTIPGPRRGSGPEITFDPLPEAAAPAGRPSASGRHKKRSR 63  
Db 854 SAHQTRSTTTPRCSPAREVRP-----APR-----PRPTASAPAAIPA----- 894  
QY 64 RVLVPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAPLPPEDAPNAASLAPT 123  
Db 895 --LVPVPVSAVPVISAPNP-----ITILPVHIL-----PSPPPPSQIPPCSSPACT 938  
QY 124 PVSPVLEPFLNTSEPSDYAL 143  
Db 939 PPPACTPPPAHTPPPAQTCL 958  
RESULT 4  
Q9Y5L9 PRELIMINARY; PRT: 2971 AA.  
AC Q9Y5L9;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)  
DE TRANSCRIPTIONAL ACTIVATOR SRCAP.  
GN SRCAP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99278407; PubMed=10347196;  
RA JOHNSTON H., Kneer J., Chackalaparampil I., Yaciuk P., Chrivia J.;  
RT "Identification of a novel SNF2/SWI2 protein family member, SRCAP,  
RT which interacts with CREB-binding protein."  
RL J. Biol. Chem. 274:16370-16376(1999).  
DR EMBL; AF143946; AAD39760.1; -  
DR INTERPRO; IPR000330; -  
DR INTERPRO; IPR000637; -  
DR INTERPRO; IPR001650; -  
DR INTERPRO; IPR002965; -  
DR PFAM; PF00176; SNF2\_N; 1.  
DR PFAM; PF00271; helicase\_C; 1.  
DR PRINTS; PRO0929; ATHOOK.  
DR PRINTS; PRO1217; PRICHEXTENS.  
SQ SEQUENCE 2971 AA; 315641 MW; F9F7EE70304B78A3 CRC64;

## Query Match

Best Local Similarity 13.9%; Score 114.5; DB 4; Length 2971;  
Matches 41; Conservative 12; Mismatches 12; Indels 37; Gaps 6;  
QY 6 SCHPTMTILQAP--TPAPSTIPGPRRGSGPEITFDPLPEAAAPAGRPSASGRHKKRSR 63  
Db 2157 SAHQTRSTTTPRCSPAREVRP-----APR-----PRPTASAPAAIPA----- 2197  
QY 64 RVLVPRVVRQLPVEEPNPAKRLFLLLTIVFCQILMAEEGVPAPLPPEDAPNAASLAPT 123  
Db 2198 --LVPVPVSAVPVISAPNP-----ITILPVHIL-----PSPPPPSQIPPCSSPACT 2241  
QY 124 PVSPVLEPFLNTSEPSDYAL 143  
Db 2242 PPPACTPPPAHTPPPAQTCL 2261  
RESULT 5  
Q65553 PRELIMINARY; PRT: 3247 AA.  
AC Q65553;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 22:04:55 ; Search time 64.84 Seconds  
(without alignments)  
281.993 Million cell updates/sec

Title: US-08-799-910-10  
Perfect score: 823  
Sequence: 1 MCHSRCHPTWTILQAPTPA.....EPSDYALDSTFLQHPAAF 156

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	771.5	93.7	193	4	O75353 homo sapien
2	122.5	14.9	423	4	O9NXH8 homo sapien
3	114.5	13.9	1668	4	O15026 homo sapien
4	114.5	13.9	2971	4	O9Y5L9 homo sapien
5	114	13.9	3247	12	Q65553 bovine herp
6	112.5	13.7	649	5	O9V884 drosophila
7	110.5	13.4	1011	3	O9P944 pneumocysti
8	109.5	13.3	552	10	Q96343 brassica na
9	109.5	13.3	1315	10	Q9SPM0 zea mays (m
10	109	13.2	285	3	P78977 yarrowia li
11	109	13.2	1709	4	O15047 homo sapien
12	109	13.2	3325	12	O9IBT9 turkey herp
13	108.5	13.2	233	2	O9L568 streptococc
14	107.5	13.1	203	2	O9RCX9 streptomyc
15	107.5	13.1	581	5	Q20517 caenorhabdi
16	107.5	13.1	1013	5	O9VYT7 drosophila
17	107	13.0	426	2	O9LAY5 streptococc
18	107	13.0	496	12	O98457 paramecium
19	106.5	12.9	901	2	Q44562 actinomyc

20	106.5	12.9	1498	4	O9Y4F5
21	106	12.9	1029	12	O9YMX0
22	106	12.9	2066	12	P89920 oat blue dw
23	105.5	12.8	249	2	O9L570 streptococc
24	105	12.8	296	5	O9V729 drosophila
25	105	12.8	351	10	Q39492 chlamydomon
26	105	12.8	549	12	O89370 paramecium
27	104.5	12.7	243	2	O9L567 streptococc
28	104.5	12.7	244	2	O9L565 streptococc
29	104.5	12.7	535	2	P74745 synchocyst
30	104.5	12.7	616	12	Q96716 chlorella v
31	104	12.6	395	2	O9LAY2 streptococc
32	104	12.6	408	2	O9LAY0 streptococc
33	103	12.5	241	5	O45522 caenorhabdi
34	102.5	12.5	288	12	O84565 paramecium
35	102.5	12.5	767	5	O62537 drosophila
36	102	12.4	236	2	O9L569 streptococc
37	102	12.4	331	5	O9V728 drosophila
38	102	12.4	708	10	O9SX31 arabidopsis
39	101.5	12.3	344	11	O89037 rattus norv
40	101	12.3	339	12	O84465 paramecium
41	101	12.3	395	2	O9LAZ1 streptococc
42	101	12.3	437	2	O9LAY4 streptococc
43	101	12.3	739	2	O9ROT4 streptococc
44	101	12.3	852	4	O9NP71 homo sapien
45	100	12.2	802	11	P70433 mus musculu

## ALIGNMENTS

RESULT 1  
O75353 PRELIMINARY; PRT; 193 AA.  
AC O75353;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
DE ANTI-DEATH PROTEIN.  
GN IEX-1L.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98369175; PubMed=9703517;  
RA Wu M.X., Ao Z., Prasad K.V.S., Wu R., Schlossman S.F.;  
RT "IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival";  
RL Science 281:998-1001(1998).  
DR EMBL; AF039067; AAC32558.1; -;  
DR EMBL; AF071596; AAC72344.1; -;  
SQ SEQUENCE 193 AA; 21028 MW; 7927D9D3FFBC7C57 CRC64;

Query Match 93.7%; Score 771.5; DB 4; Length 193;  
Best Local Similarity 78.8%; Pred. No. 8.1e-64;  
Matches 152; Conservative 1; Mismatches 3; Indels 37; Gaps 1;

Oy	1	MCHSRCHPTWTILQAPTPAPSTIPGPRRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK	60
Db	1	MCHSRCHPTWTILQAPTPAPSTIPGPRRGSGPEIFTFDPLPEPAAAPAGRPSASRGHRK	60
Oy	61	RSRRVLPYPR-----VVRRQLPVEENPA	83
Db	61	RSRRVLPYPR-----VVRRQLPVEENPA	120
Oy	84	KRLFLLLTIVFCOILMAEGVPAIPPEADPNAAISLAPTPSVPLPGPNLTSPSDYAL	143
Db	121	KRLFLLLTIVFCOILMAEGVPAIPPEADPNAAISLAPTPSVPLPGPNLTSPSDYAL	180
Oy	144	DLSTFLQHPAAF	156

---

QY 21 LTIVFCQILMA-----EEGVPAFLPPEDAPNAASLAPTPVSPVLEPNLTSEP-SDYAL 73  
db 793 ittvscgkmaaptrntcttgidppppgspnitsvshsvkvkfsfgeahgpkayav 852  
QY 74 DLSTFLQOHPPAA 85  
Db 853 ilttgeahpsa 864

RESULT 15  
W37148  
ID W37148 standard; Protein; 541 AA.  
XX AC W37148;  
XX 06-JUL-1998 (first entry)  
XX Mammalian Ena (Mena).  
XX Mena protein; mammalian Ena; Enabled; Evi protein; cytoskeleton;  
KW cell morphology; cell adhesion; cell differentiation; cell growth;  
KW cell motility; mouse.  
XX Mus musculus.

Key	Location/Qualifiers
Domain	1..113
Peptide	/note="Ena-VASP homology domain 1 (EVH1)"
Peptide	156..160
Peptide	/note="LERER repeat"
Peptide	166..170
Peptide	/note="LERER repeat"
Peptide	171..175
Peptide	/note="LERER repeat"
Peptide	176..180
Peptide	/note="LERER repeat"
Peptide	199..203
Peptide	/note="LERER repeat"
Peptide	207..211
Peptide	/note="LERER repeat"
Modified-site	236
Binding-site	/note="O-phosphorylated"
Binding-site	305..310
Binding-site	/note="mediates binding to profilin"
Binding-site	316..321
Binding-site	/note="mediates binding to profilin"

WO9801755-A1.  
15-JAN-1998.  
03-JUL-1997; 97WO-US11669.  
05-JUL-1996; 96US-0675815.  
(GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.  
(HUTC-) HUTCHINSON CANCER RES CENT FRED.  
Certlier FB, Niebuhr K, Soriano P, Wehland J;  
WPI: 1998-101197/09.  
DR N-PSDB; V02996.  
XX Detection of modulators of Mena and Ena-VASP-like genes and proteins  
PT - used in control of cytoskeletal dynamic events in normal and  
PT abnormal cell morphology, adhesion, motility, growth and  
PT differentiation  
XX  
PS Example 1; Page 54-56; 77pp; English.  
XX  
CC This polypeptide comprises murine Mena (mammalian Ena) that shows  
CC significant amino acid similarity to Drosophila Ena and which  
CC exhibits a broad pattern of tissue distribution in neurons,

CC fibroblasts, kidney epithelium, muscle, neural crest and  
CC haematopoietic cells. Its amino acid sequence was deduced from  
CC a cDNA clone (see V02996) obtained from a mouse embryonic stem cell  
CC cDNA library. 2 Novel mammalian genes, Mena and Ena-VASP-like  
CC (Evi, see V02997), encoding novel proteins Mena and Evi (see  
CC W37148) are disclosed. Mena and Evi proteins have a discrete, EVH1  
CC functional domain responsible for Mena binding to Listeria, and to  
CC the cytoskeletal proteins zyxin and vinculin. The EVH1 domain of  
CC Mena is also responsible and sufficient for targeting localisation  
CC of Mena and Mena-based fusion proteins to focal adhesions, and to  
CC the surface of Listeria cells at the polar site of induction of  
CC actin comet tail formation. Based on the disclosed Mena and Evi  
CC genes and proteins, a variety of methods and compositions are  
CC provided for screening, isolating and characterising endogenous and  
CC exogenous factors, drugs and therapeutic agents useful to evaluate  
CC and/or control cytoskeletal dynamic events involved in normal and  
CC abnormal cell morphology, adhesion, motility, growth and/or  
CC differentiation. A method of detecting a modulator of Mena  
CC activity/expression is claimed.  
XX  
SQ Sequence 541 AA;

Query Match 15.28; Score 67; DB 19; Length 541;  
Best Local Similarity 29.9%; Pred. No. 25;  
Matches 20; Conservative 4; Mismatches 23; Indels 20; Gaps 2;  
QY 6 PVSEPNPAKRLLLFLLTIVFCQILMAEEGVPAFLPPEDAPN-----AASLAPTPVSPVLEP 61  
Db 268 paesptpqgvl-----gpppppppplpsgpayasalpppppppppp 311  
QY 62 FNLTSEP 68  
Db 312 lpstgpp 318

Search completed: January 30, 2001, 22:05:37  
Job time: 11886 sec







DT 08-AUG-2000 (first entry)  
XX Human PRO1431 (UNQ737) amino acid sequence SEQ ID NO:315.  
DE  
XX  
XX  
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;  
transmembrane; secretion; immunoadhesion; pharmaceutical; screening.  
XX  
OS Homo sapiens.  
PN WO200012708-A2.  
XX  
XX 09-MAR-2000.  
XX  
XX 01-SEP-1999; 99WO-US20111.  
XX 01-SEP-1998; 98US-0098716.  
PR 01-SEP-1998; 98US-0098749.  
PR 01-SEP-1998; 98US-0098750.  
PR 02-SEP-1998; 98US-0098803.  
PR 02-SEP-1998; 98US-0098821.  
PR 02-SEP-1998; 98US-0098843.  
PR 09-SEP-1998; 98US-0099536.  
PR 09-SEP-1998; 98US-0099596.  
PR 09-SEP-1998; 98US-0099598.  
PR 09-SEP-1998; 98US-0099602.  
PR 09-SEP-1998; 98US-0099642.  
PR 10-SEP-1998; 98US-0099741.  
PR 10-SEP-1998; 98US-0099754.  
PR 10-SEP-1998; 98US-0099763.  
PR 10-SEP-1998; 98US-0099792.  
PR 10-SEP-1998; 98US-0099808.  
PR 10-SEP-1998; 98US-0099812.  
PR 10-SEP-1998; 98US-0099815.  
PR 10-SEP-1998; 98US-0099816.  
PR 15-SEP-1998; 98US-0100385.  
PR 15-SEP-1998; 98US-0100388.  
PR 15-SEP-1998; 98US-0100390.  
PR 16-SEP-1998; 98US-0100384.  
PR 16-SEP-1998; 98US-0100627.  
PR 16-SEP-1998; 98US-0100661.  
PR 16-SEP-1998; 98US-0100662.  
PR 16-SEP-1998; 98US-0100664.  
PR 17-SEP-1998; 98US-0100683.  
PR 17-SEP-1998; 98US-0100684.  
PR 17-SEP-1998; 98US-0100710.  
PR 17-SEP-1998; 98US-0100711.  
PR 17-SEP-1998; 98US-0100919.  
PR 17-SEP-1998; 98US-0100930.  
PR 18-SEP-1998; 98US-0100848.  
PR 18-SEP-1998; 98US-0100849.  
PR 18-SEP-1998; 98US-0101014.  
PR 18-SEP-1998; 98US-0101068.  
PR 18-SEP-1998; 98US-0101071.  
PR 22-SEP-1998; 98US-0101279.  
PR 23-SEP-1998; 98US-0101471.  
PR 23-SEP-1998; 98US-0101472.  
PR 23-SEP-1998; 98US-0101474.  
PR 23-SEP-1998; 98US-0101475.  
PR 23-SEP-1998; 98US-0101476.  
PR 23-SEP-1998; 98US-0101477.  
PR 24-SEP-1998; 98US-0101479.  
PR 24-SEP-1998; 98US-0101738.  
PR 24-SEP-1998; 98US-0101741.  
PR 24-SEP-1998; 98US-0101743.  
PR 24-SEP-1998; 98US-0101915.  
PR 24-SEP-1998; 98US-0101916.  
PR 29-SEP-1998; 98US-0102207.  
PR 29-SEP-1998; 98US-0102240.  
PR 29-SEP-1998; 98US-0102307.  
PR 29-SEP-1998; 98US-0102330.  
PR 29-SEP-1998; 98US-0102331.  
PR 30-SEP-1998; 98US-0102484.  
PR 30-SEP-1998; 98US-0102487.

PR 30-SEP-1998; 98US-0102570.  
PR 30-SEP-1998; 98US-0102571.  
PR 01-OCT-1998; 98US-0102684.  
PR 01-OCT-1998; 98US-0102687.  
PR 02-OCT-1998; 98US-0102965.  
PR 06-OCT-1998; 98US-0103258.  
PR 06-OCT-1998; 98US-0103449.  
PR 07-OCT-1998; 98US-0103314.  
PR 07-OCT-1998; 98US-0103315.  
PR 07-OCT-1998; 98US-0103328.  
PR 07-OCT-1998; 98US-0103395.  
PR 07-OCT-1998; 98US-0103396.  
PR 07-OCT-1998; 98US-0103401.  
PR 08-OCT-1998; 98US-0103633.  
PR 08-OCT-1998; 98US-0103678.  
PR 08-OCT-1998; 98US-0103679.  
PR 08-OCT-1998; 98US-0103711.  
PR 14-OCT-1998; 98US-0104257.  
PR 20-OCT-1998; 98US-0104987.  
PR 20-OCT-1998; 98US-0105000.  
PR 21-OCT-1998; 98US-0105104.  
PR 22-OCT-1998; 98US-0105169.  
PR 22-OCT-1998; 98US-0105266.  
PR 26-OCT-1998; 98US-0105693.  
PR 26-OCT-1998; 98US-0105694.  
PR 27-OCT-1998; 98US-0105807.  
PR 27-OCT-1998; 98US-0105881.  
PR 27-OCT-1998; 98US-0105882.  
PR 27-OCT-1998; 98US-0106062.  
PR 28-OCT-1998; 98US-0106023.  
PR 28-OCT-1998; 98US-0106029.  
PR 28-OCT-1998; 98US-0106030.  
PR 28-OCT-1998; 98US-0106032.  
PR 28-OCT-1998; 98US-0106033.  
PR 28-OCT-1998; 98US-0106178.  
PR 29-OCT-1998; 98US-0106248.  
PR 29-OCT-1998; 98US-0106384.  
PR 30-OCT-1998; 98US-0108500.  
PR 30-OCT-1998; 98US-0106464.  
PR 03-NOV-1998; 98US-0106856.  
PR 03-NOV-1998; 98US-0108902.  
PR 03-NOV-1998; 98US-0108905.  
PR 03-NOV-1998; 98US-0108919.  
PR 03-NOV-1998; 98US-0106932.  
PR 03-NOV-1998; 98US-0106934.  
PR 03-NOV-1998; 98US-0107783.  
PR 17-NOV-1998; 98US-0108775.  
PR 17-NOV-1998; 98US-0108779.  
PR 17-NOV-1998; 98US-0108787.  
PR 17-NOV-1998; 98US-0108788.  
PR 17-NOV-1998; 98US-0108801.  
PR 17-NOV-1998; 98US-0108802.  
PR 17-NOV-1998; 98US-0108806.  
PR 17-NOV-1998; 98US-0108807.  
PR 17-NOV-1998; 98US-0108867.  
PR 17-NOV-1998; 98US-0108925.  
PR 18-NOV-1998; 98US-0108848.  
PR 18-NOV-1998; 98US-0108849.  
PR 18-NOV-1998; 98US-0108850.  
PR 18-NOV-1998; 98US-0108851.  
PR 18-NOV-1998; 98US-0108852.  
PR 18-NOV-1998; 98US-0108858.  
PR 18-NOV-1998; 98US-0108904.  
XX  
XX (GETH ) GENENTECH INC.  
PA  
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;  
PI WPI: 2000-237871/20.  
XX N-PSDB: A37110.  
DR  
DR New mammalian DNA sequences encoding transmembrane, receptor or  
XX  
PT



W18201  
XX W18201 standard; protein; 610 AA.  
XX W18201;  
XX W18201;  
XX 18-AUG-1997 (first entry)  
XX Platelet glycoprotein Ib alpha naturally-occurring wild-type.  
DE  
XX GPIb alpha; recombinant polypeptide; antithrombotic;  
XX platelet adhesion; platelet aggregation; thrombus formation;  
KW platelet-type von Willebrand disease.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH 36..200  
FT /label= Leucine\_rich  
FT 220..310  
FT /label= Hinge  
FT /note= "A major binding site for von Willebrand factor"  
FT 228..238  
FT /note= "Preferred sites for mutations that result  
FT in a mutant polypeptide having more reactivity  
FT with von Willebrand factor"  
FT  
FT Misc-difference 233  
FT /note= "Preferred mutation; substitution of Val for Gly"  
FT Region 310..420  
FT /label= Serine/threonine\_rich  
FT  
XX US5624817-A.  
XX  
XX 29-APR-1997.  
XX  
XX 07-OCT-1991; 91US-0770968.  
XX  
XX 07-OCT-1991; 91US-0770968.  
PR 28-APR-1994; 94US-0234265.  
XX  
XX (UUNY ) UNIV NEW YORK STATE RES FOUND.  
XX  
XX Cunningham D, Finch CN, Lyle VA, Miller JL, Pincus MR;  
PI WPI; 1997-258215/23.  
XX  
XX DNA encoding mutant platelet glyco:protein Ib alpha - for production  
PT of recombinant polypeptide(s) useful as antithrombotic agents, etc.  
XX  
XX Disclosure; Column 27-32; 20pp; English.  
XX  
XX The present sequence represents the naturally-occurring wild-type  
CC platelet glycoprotein Ib alpha (GPIb alpha). The polypeptide can have a  
CC mutation which makes it more reactive with von Willebrand factor (vWF)  
CC than the wild-type. Preferably the mutation occurs in the hinge region of  
CC within residues 228 to 238. The preferred mutation is a substitution of  
CC Val for the wild-type Gly at position 233. The mutation alters the  
CC three-dimensional structure of the mutant polypeptide from a beta-bend  
CC conformation to an alpha helix conformation, and also creates an  
CC amphipathic region within the polypeptide. The DNA encoding mutant  
CC platelet GPIb alpha polypeptides (where the mutation is between the  
CC residues 228 to 238, but not at position 233), vectors and cells are  
CC useful for production of recombinant mutant GPIb alpha polypeptides,  
CC which can be used to inhibit platelet adhesion and aggregation (e.g. in  
CC antithrombotic medicaments or on medical devices or vascular  
CC prostheses), or can be labelled and used as imaging agents, or can be  
CC coupled to a thrombolytic agent for targeting to sites of thrombus  
CC formation. The oligonucleotides can be used as probes for diagnosis of  
CC platelet-type von Willebrand disease.  
XX  
XX Sequence 610 AA;  
SQ

Query Match  
16.28; Score 71.5; DB 18; Length 610;

RESULT	6
R89436	
ID	R89436 standard; Protein; 610 AA.
XX	
AC	R89436;
XX	
DT	02-SEP-1996 (first entry)
XX	
DE	Mutated platelet glycoprotein-Ib-alpha GPIb protein sequence.
XX	
KW	Platelet glycoprotein-Ib-alpha; GPIb; mutagenesis; point mutation;
KW	von Willebrand factor; blood disorder; platelet disorder;
KW	protein engineering; Bernard-Soulier disease.
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 57
FT	/note= "substitution from Phe in wild-typ
FT	GPIb-alpha"
FT	Region 36..200
FT	/note= "Leu rich area"
FT	Region 200...220
FT	/note= "flanking region to Leu rich area"
FT	Region 220..310
FT	/note= "hinge region"
FT	Region 310..420
FT	/note= "Ser/Thr rich area"
XX	
PN	US5492809-A.
XX	
PD	20-FEB-1996.
XX	
PF	07-OCT-1991; 91US-0770968.
XX	
PR	15-JAN-1992; 92US-0821717.
PR	07-OCT-1991; 91US-0770968.
PR	09-SEP-1993; 93US-0119262.
XX	
PA	(UUNY ) UNIV NEW YORK STATE RES FOUND.
XX	
PI	Cunningham D, Finch CN, Lyle VA, Miller JL;
XX	
DR	WPI; 1996-128585/13.
XX	
PT	DNA encoding platelet glyco:protein Ib alpha mutant phe57 -
PT	introduced into platelets to reduce aggregation and reactivity with
PT	von Willebrand factor, also probe for diagnosis of Bernard-Soulier
PT	disease
XX	
PS	Disclosure; Column 21-26; 20pp; English.
XX	
CC	A substitution of T for C at position 259 in the DNA sequence of
CC	GPIb-alpha leads to the replacement of Phe for Leu at
CC	residue 57 of the mature GPIb-alpha molecule. This mutated GPIb-
CC	alpha protein is less reactive with von Willebrand factor, a
CC	characteristic of the autosomal recessive bleeding disorder,
CC	Bernard-Soulier disease.
XX	
SO	Sequence 610 AA;

Query Match 16.2%; Score 71.5; DB 17; Length 610;  
 Best Local Similarity 45.0%; Pred. No. 9.5;  
 Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;

Oy 36 PAPLP-----PEDAPNAAASLAPTVPSPVLEPFFNLTSPPS 69  
 | | | | | | | | | | | | | | | | | | | |  
 Db 366 pfpspttsepyvpepapnttletpstpttep---tsespa 402

## RESULT 7

CC three-dimensional coordinates obtained from the crystals, is useful  
 CC for identifying an agent that stabilizes the Ras-Sos complex. The  
 CC crystals are also useful for identifying agents that inhibit the  
 CC formation of Ras-Sos complex. Ras and Sos fragments are useful for  
 CC growing a crystal of a protein-ligand complex. Agents that stabilize  
 CC or inhibit the formation of Ras-Sos complex are useful in the  
 CC treatment of cancer. The present sequence represents a Drosophila  
 CC Sos protien.

XX SQ Sequence 1596 AA;

Query Match 18.0%; Score 79.5; DB 21; Length 1596;  
 Best Local Similarity 28.3%; Pred. No. 4.2;  
 Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps 6;

OY 1 VRRQLPVE-----EPNPAKRLLLFTLVFCQILMAEKGVPAPLPEDAPNAASLAPT 53  
 Db 1423 irnsaiekraaatsqnaagpisttlvtvsqavatdeplplisp-----aasstt 1477  
 OY 54 --PVSPVLEPF--NLTSEP-----SDYALDLSTFLOQ-----HPAAF 86  
 Db 1478 tspltpamspnlpshpvestsssyahqlrmrqqqqqqthpaly 1523

RESULT 4

RS1116 ID R51116 standard; protein; 610 AA.

XX AC R51116;

DT 23-SEP-1994 (first entry)

XX DE Platelet glycoprotein Ib alpha.

XX KW Platelet; glycoprotein; imaging; thrombolytic agent;  
 KW tissue plasminogen activator; tPA; pro-urokinase; urokinase;  
 KW streptokinase; Bernard-Soulier disease; thrombus; aggregation;  
 KW anisoylated plasminogen-streptokinase activator complex; adhesion;  
 KW inhibition.

XX OS Homo sapiens.

XX US298239-A.

XX PD 29-MAR-1994.

XX PF 07-OCT-1991; 91US-0770968.

XX PR 07-OCT-1991; 91US-0770968.

XX PR 15-JAN-1992; 92US-0821717.

XX PA (UUNY ) UNIV NEW YORK STATE RES FOUND.

XX PI Cunningham D, Finch CN, Lyle VA, Miller JL;

XX DR WPI; 1994-100287/12.

XX PT Platelet glycoprotein Ib alpha with an amino acid substn at  
 PT position 57 - has reduced reactivity with Von Willebrand factor,  
 PT and can be used to inhibit platelet aggregation and inhibition

XX PS Claim 1; Columns 3-8; 20pp; English.

XX CC A substitution in platelet glycoprotein Ib alpha (Leucine 57 to  
 CC Phenylalanine) underlies a form of Bernard-Soulier disease. The  
 CC mutated glycoprotein can be used in compositions to inhibit  
 CC platelet aggregation/adhesion. The glycoprotein may be labelled and  
 CC used as an imaging agent and may also be bound to a thrombolytic  
 CC agent, preferably tissue plasminogen activator (tPA),  
 CC (pro) urokinase, streptokinase, anisoylated plasminogen-streptokinase  
 CC activator complex, tPA analogues or a protease, allowing localisation  
 CC of the thrombolytic agent to a thrombus.

XX SQ Sequence 610 AA;

Query Match 16.2%; Score 71.5; DB 15; Length 610;  
 Best Local Similarity 45.0%; Pred. No. 9.5;  
 Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;

OY 36 PAPLP-----PEDAPNAASLAPTPVSPVLEPFLNLTSEPS 69  
 Db 366 ptpspstsepyepapapmttleptpspttpep----tsepa 402

RESULT 5

RS5664 ID R5664 standard; protein; 610 AA.

XX AC R5664;

DT 23-SEP-1994 (first entry)

XX DE Mutant platelet glycoprotein Ib alpha.

XX KW Platelet; glycoprotein; imaging; thrombolytic agent;  
 KW tissue plasminogen activator; tPA; pro-urokinase; urokinase;  
 KW streptokinase; Bernard-Soulier disease; thrombus; aggregation;  
 KW anisoylated plasminogen-streptokinase activator complex; adhesion;  
 KW inhibition.

XX OS Homo sapiens.

XX PN US298239-A.

XX PD 29-MAR-1994.

XX PF 07-OCT-1991; 91US-0770968.

XX PR 07-OCT-1991; 91US-0770968.

XX PR 15-JAN-1992; 92US-0821717.

XX PA (UUNY ) UNIV NEW YORK STATE RES FOUND.

XX PI Cunningham D, Finch CN, Lyle VA, Miller JL;

XX DR WPI; 1994-100287/12.

XX PT Platelet glycoprotein Ib alpha with an amino acid substn at  
 PT position 57 - has reduced reactivity with Von Willebrand factor,  
 PT and can be used to inhibit platelet aggregation and inhibition

XX PS Claim 1; 20pp; English.

XX CC A substitution in platelet glycoprotein Ib alpha (Leucine 57 to  
 CC Phenylalanine) underlies a form of Bernard-Soulier disease. The  
 CC mutated glycoprotein can be used in compositions to inhibit  
 CC platelet aggregation/adhesion. The glycoprotein may be labelled and  
 CC used as an imaging agent and may also be bound to a thrombolytic  
 CC agent, preferably tissue plasminogen activator (tPA),  
 CC (pro) urokinase, streptokinase, anisoylated plasminogen-streptokinase  
 CC activator complex, tPA analogues or a protease, allowing localisation  
 CC of the thrombolytic agent to a thrombus.

XX SQ Sequence 610 AA;

Query Match 16.2%; Score 71.5; DB 15; Length 610;  
 Best Local Similarity 45.0%; Pred. No. 9.5;  
 Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;

OY 36 PAPLP-----PEDAPNAASLAPTPVSPVLEPFLNLTSEPS 69  
 Db 366 ptpspstsepyepapapmttleptpspttpep----tsepa 402

PT disease, e.g. atherosclerosis, restenosis, hypertension, etc  
 XX  
 PS Example 6; Fig 5; 163pp; English.  
 XX  
 CC This protein is encoded by the novel human fchd605 gene (see  
 CC T94471) that is up-regulated in monocytes treated with oxidised  
 CC low density lipoproteins that simulate the conditions under which  
 CC foam cells develop during atherogenesis. The protein has sequence  
 CC similarity to the mouse gly96 gene and to EST T49532. Novel  
 CC fchd531, fchd540, fchd545, fchd602 and fchd605 genes (see T94467-71)  
 CC provide a fingerprint for the study of cardiovascular diseases,  
 CC including atherosclerosis, ischaemia/reperfusion, hypertension,  
 CC restenosis and arterial inflammation. Methods are provided for the  
 CC diagnosis, monitoring in clinical trials, screening for  
 CC therapeutically effective compounds, and treatment of  
 CC cardiovascular diseases based on discoveries regarding the  
 CC expression patterns of these novel genes.  
 XX  
 SQ Sequence 156 AA:

Query Match 100.0%; Score 442; DB 18; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-39;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VRRQLPVEPNPAKRLLLFTVFCQILMAEEGVPAPLPPEDAPNAASLAPTVPSPVLE 60  
 DB 71 VRRQLPVEPNPAKRLLLFTVFCQILMAEEGVPAPLPPEDAPNAASLAPTVPSPVLE 130

OY 61 PFNLTSEPSDYALDLSTFLQHPAAF 86  
 DB 131 PFNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 2  
 Y45017  
 ID Y45017 standard; Protein: 156 AA.  
 AC Y45017:  
 DT 31-MAY-2000 (first entry)  
 XX  
 DE Protein encoded by fchd605 gene.  
 KW fchd605 gene; human; cardiovascular disease; oncogenic disorder;  
 KW diabetic retinopathy; fibroproliferative disorder; atherosclerosis;  
 KW TGF-beta signalling pathway; TGF; Transforming growth factor;  
 KW pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;  
 KW vascularisation; cytostatic; antidiabetic; ophthalmological.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200006206-A1.  
 XX  
 PD 10-FEB-2000.  
 XX  
 PF 30-JUL-1999; 99WO-US17394.  
 XX  
 PR 30-JUL-1998; 98US-0126640.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Falb DA:  
 XX  
 DR WPI: 2000-205414/18.  
 DR N-PSDB; 250711.  
 XX  
 PT Identifying substances for ameliorating symptoms of fibroproliferative  
 PT diseases or oncogenic related disorders.  
 XX  
 PS Example; Fig 5; 214pp; English.  
 XX  
 CC The patent discloses methods for the treatment and diagnosis of

CC cardiovascular diseases by novel human genes which are differentially  
 CC expressed in different cardiovascular disease states. Compositions which  
 CC can modify TGF-beta signalling pathway are identified by screening.  
 CC These are used therapeutically to treat fibroproliferative and oncogenic  
 CC disorders, especially tcf (transforming growth factor)-beta related  
 CC disorders, including diabetic retinopathy, atherosclerosis, pancreatic  
 CC cancer, angiogenesis, inflammation, fibrosis, tumour growth and  
 CC vascularisation. The present sequence is the protein product of fchd605  
 CC gene which is up-regulated in monocytes treated with oxidised LDL (low  
 CC density lipoprotein) can be used to design cardiovascular disease  
 CC treatment strategies. Depending on whether the up-regulation has a  
 CC pathogenic or protective effect treatment methods can be designed to  
 CC increase or decrease the activity of the protein product of the gene.  
 XX  
 SQ Sequence 156 AA:

Query Match 100.0%; Score 442; DB 21; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-39;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VRRQLPVEPNPAKRLLLFTVFCQILMAEEGVPAPLPPEDAPNAASLAPTVPSPVLE 60  
 DB 71 VRRQLPVEPNPAKRLLLFTVFCQILMAEEGVPAPLPPEDAPNAASLAPTVPSPVLE 130

OY 61 PFNLTSEPSDYALDLSTFLQHPAAF 86  
 DB 131 PFNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 3  
 Y68821  
 ID Y68821 standard; protein; 1596 AA.  
 AC Y68821:  
 DT 16-MAY-2000 (first entry)  
 XX  
 DE Amino acid sequence of a Drosophila Son of sevenless (Sos) protein.  
 KW Ras; Son of sevenless; Sos; crystal; Ras-Sos complex; cancer;  
 KW protein coordinate data.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200005258-A1.  
 XX  
 PD 03-FEB-2000.  
 XX  
 PF 20-JUL-1999; 99WO-US16348.  
 XX  
 PR 21-JUL-1998; 98US-0119794.  
 XX  
 PA (UYRQ ) UNIV ROCKEFELLER.  
 XX  
 PI Boriack-Sjodin A, Margarit SM, Bar-Sagi D, Cole P, Kuriyan J;  
 XX WPI: 2000-182647/16.  
 XX  
 PT Novel crystals comprising a Ras-Son of sevenless complex, useful for  
 PT screening drugs useful in cancer treatment  
 XX  
 PS Disclosure; Page 169-175; 224pp; English.  
 XX  
 CC The specification describes a crystal complex comprising at least a  
 CC Sos contacting region of a Ras protein and at least a Son of sevenless  
 CC (Sos) protein catalytic region fragment, that effectively diffracts  
 CC x-rays. Ras and Sos form a tight complex. Sos does not impede the  
 CC binding sites for the nucleotide base and the ribose of GTP or GDP  
 CC and thus the Ras-Sos complex maintains a structure that permits  
 CC nucleotide release and rebinding. The crystals are used for the  
 CC determination of the atomic coordinates of the complex to a resolution  
 CC of more than 5.0 Angstrom. The crystals, or a dataset comprising the

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 22:05:37 ; Search time 54.97 Seconds  
(without alignments)  
53.496 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156  
Perfect score: 442  
Sequence: 1 VRQLPVEEPNPAKRLLE.....EPSDYALDLSTFLQHPAAF 86

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: /cgn2\_2/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /cgn2\_2/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /cgn2\_2/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /cgn2\_2/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /cgn2\_2/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /cgn2\_2/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /cgn2\_2/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /cgn2\_2/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /cgn2\_2/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /cgn2\_2/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /cgn2\_2/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /cgn2\_2/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /cgn2\_2/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /cgn2\_2/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /cgn2\_2/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /cgn2\_2/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /cgn2\_2/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /cgn2\_2/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /cgn2\_2/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /cgn2\_2/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /cgn2\_2/gcgdata/geneseq/geneseq/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	442	100.0	156	18	W36006 Human Fchd605 gene
2	442	100.0	156	21	Y45017 Protein encoded by
3	79.5	18.0	1596	21	Y68821 Amino acid sequenc
4	71.5	16.2	610	15	R31116 Platelet glycoprot
5	71.5	16.2	610	15	R36664 Mutant platelet g
6	71.5	16.2	610	17	R89436 Mutated platelet g
7	71.5	16.2	610	18	W18201 platelet glycoprot
8	69.5	15.7	739	18	W25790 Gene 036 product d
9	68.5	15.5	204	18	W14578 Streptococcus pneu
10	68.5	15.5	370	21	Y99428 Human PRO1431 (UNQ
11	68.5	15.5	878	15	R55060 Sequence of human
12	68.5	15.5	878	16	R85487 Human E-cadherin p

13	68.5	15.5	882	20	Y09375 Wild-type E-cadher
14	68.5	15.5	1337	16	R85203 huDEP-1. Homo sap
15	67	15.2	541	19	W37148 Mammalian Ena (Men
16	67	15.2	783	19	W37151 Mouse neural Mena+
17	67	15.2	787	19	W37152 Mouse neural Mena+
18	67	15.2	802	19	W37153 Mouse neural Mena+
19	66.5	15.0	131	18	W27646 Secreted protein A
20	66.5	15.0	131	18	W4082 Human secreted pro
21	65.5	14.8	128	18	W14577 Streptococcus pneu
22	65.5	14.8	286	17	W03566 Mycobacterium tube
23	65.5	14.8	325	17	W03565 Mycobacterium tube
24	65.5	14.8	332	18	W32418 Mycobacterium tube
25	65.5	14.8	332	18	W32350 Mycobacterium tube
26	65.5	14.8	332	19	W81683 M. tuberculosis im
27	65.5	14.8	332	19	W64322 Mycobacterium tube
28	65.5	14.8	332	20	Y39083 M. tuberculosis an
29	65.5	14.8	332	20	Y38945 M. tuberculosis re
30	65.5	14.8	456	20	Y17067 Human 3-OST-4 prot
31	65.5	14.8	552	20	Y39225 M. tuberculosis fu
32	65.5	14.8	552	20	Y39082 M. tuberculosis fu
33	65.5	14.8	802	19	W81746 M. tuberculosis fu
34	65.5	14.8	802	19	W64379 Mycobacterium anti
35	65.5	14.8	802	20	Y32063 Mycobacterium tube
36	65.5	14.8	802	20	Y39224 Mycobacterium fu
37	65.5	14.8	802	20	Y39176 M. tuberculosis fu
38	65.5	14.8	802	20	Y39081 M. tuberculosis fu
39	65.5	14.8	802	20	Y39033 M. tuberculosis fu
40	65.5	14.8	1093	16	R66460 AF-17 protein. Ho
41	65	14.7	328	19	W36984 E. coli ZipA prote
42	65	14.7	427	21	Y68748 Amino acid sequenc
43	64.5	14.6	147	15	R60799 Rape abscission/de
44	64.5	14.6	194	18	W14584 Streptococcus pneu
45	64.5	14.6	412	20	Y49151 Amino acid sequenc

ALIGNMENTS

RESULT	1
W36006	
ID	W36006 standard; Protein: 156 AA.
XX	
AC	W36006;
XX	
DT	03-MAR-1998 (first entry)
DE	Human Fchd605 gene product.
XX	
XX	Fchd605 gene; differential expression; monocyte; human;
KW	foam cell; cardiovascular disease; atherosclerosis; ischaemia;
KW	referrusion; hypertension; restenosis; arterial inflammation;
KW	therapy; diagnosis; drug screening; marker.
OS	Homo sapiens.
XX	
PN	W09730065-A1.
XX	
PD	21-AUG-1997.
XX	
PF	14-FEB-1997; 97WO-US02291.
XX	
PR	13-FEB-1997; 97US-0799910.
PR	16-FEB-1996; 96US-0011787.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Falb DA;
XX	
XX	WPI; 1997-424966/39.
DR	N-PSDB; T94471.
XX	
PT	New genes differentially expressed in cardiovascular disease - used for diagnosis, drug screening and treatment of cardiovascular

---

```

: ATTORNEY/AGENT INFORMATION:
:
: NAME: OBLON, NORMAN F
:
: REGISTRATION NUMBER: 24,618
:
: REFERENCE/DOCKET NUMBER: 226640720
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: 703-413-3000
:
: TELEFAX: 703-413-2220
:
: INFORMATION FOR SEQ ID NO: 3:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 286 amino acids
:
: TYPE: amino acid
:
: STRANDEDNESS:
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: peptide
:
: HYPOTHETICAL: NO
:
: US-08-382-184-3

```

Query Match 14.8%; Score 65.5; DB 1; Length 286;  
Best Local Similarity 46.9%; Pred. NO. 5;  
Matches 15: Conservative 3; Mismatches 11; Indels 3; Gaps 1;

**Qy**    29 LMAEEGVAPLPPEDAP---NAASLAPTVPSP 57  
         | : | | | | | | | : | | | | : |  
**Db**    247 LVAPPPAPAPAEPAPAPAPAGEVAPTPTTP 278

RESULT 15

US-08-641-356-3  
; Sequence 3, Application US/08641356  
; Patent No. 5866130  
; GENERAL INFORMATION:

APPLICANT:

[illegible]

## NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &  
ADDRESSEE: NEUSTADT, P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

21F. 2220Z  
COMPUTER READABLE FORM:

```

COMMITMENT TO THE COMMUNITY
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/641,356
FILING DATE:

```

CLASSIFICATION: 536

PRIOR APPLICATION DATA: US/08/382,184  
APPLICATION NUMBER:  
FILING DATE: 01-FEB-1995

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 22640720  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220

INFORMATION FOR SEQ ID NO: 3:

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 286 amino acids
; TYPE: amino acid
; STRANDEDNESS:

```

TOPOLOGY: linear

MOLECULE TYPE: peptide  
HYPOTHETICAL: NO

US-08-641-356-3

Query Match	14.88;	Score 65.5;	DB 2;	Length 280;
Best Local Similarity	46.9%;	Pred. No. 5;		
Matches	15;	Conservative	3;	Mismatches 11; Indels
Qy	29	LMAEEGVPAPLPDPDAP---	NAASLAPTPVSP	57
Db	247	LVAPPAPAPAPAPAPAPAPAGVAPTPTT	278	

Search completed: January 30, 2001, 22:06:26  
Job time: 9705 sec



```
Matches 26: Conservative 11; Mismatches 31; Indels 33; Gaps 5;
QY 1 VRRLQVPEPNPAKRLLELLLTIVFC-----QILMAEGVPAPLPDPDAPNAASLA--- 51
DB 584 VNDNAPIPEPR-----TIFFCERNRPQVINIHD---ADLPPTSPTAELTHGR 630
QY 52 -----PTVPSPVLEPFLNTSEPSDYALDLSTFLQOH 82
DB 631 VPNWTIQNDPTQESIIILKP-KMALEVGDYKINKLMDNQN 670

RESULT 12
US-08-854-585-2
; Sequence 2, Application US/08854585
; Patent No. 6114140
; GENERAL INFORMATION:
; APPLICANT: Tonks, Nicholas K. and stman, Arne
; TITLE OF INVENTION: Density Enhanced Protein Tyrosine Phosphatase
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, Suite 6300
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,585
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/237,940
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Borun, Michael F.
; REGISTRATION NUMBER: 25,447
; REFERENCE/DOCKET NUMBER: 27866/31954
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 2:
; LENGTH: 1337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US854-585-2

Query Match 15.5%; Score 68.5; DB 3; Length 1337;
Best Local Similarity 29.2%; Pred. No. 15;
Matches 21; Conservative 11; Mismatches 33; Indels 7; Gaps 2;
QY 21 LTIIVFCQILMA-----EGVPAPLPDPDAPNAASLAPTVPSPVLEPFLNTSEP-SDYAL 73
DB 793 ITTVSCGKMAPTRTCTTGITDPPDGNITSVSHNSVKVFKFSGEASHGPIKAYAV 852
QY 74 DLSTFLQOHPPAA 85
DB 853 ILTTGEAGHPSA 864

RESULT 14
US-08-382-184-3
; Sequence 3, Application US/08382184
; Patent No. 5714593
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: MICROBACTERIAL PROTEINS,
; TITLE OF INVENTION: MICROORGANISMS PRODUCING THEM AND THEIR USE FOR VACCINE
; TITLE OF INVENTION: AND FOR THE DETECTION OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER &
; ADDRESSEE: NEUSTADT, P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/382,184
; FILING DATE: 01-FEB-1995
; CLASSIFICATION: 435
```



factor to the

RESULT 8

US-08-234-265A-11  
: Sequence 11. Application US/08234265A

GENERAL INFORMATION:  
APPLICANT: Miller, Jonathan L.

APPLICANT: Cunningham, Lyle, Vicki

APPLICANT: Finch, Clara N.  
APPLICANT: Pincus, Matthew R.

APPLICANT: FALCUS/ HALLUCIN  
TITLE OF INVENTION: Mutations in the Gene Encoding the Chain of platelet Glycoprotein Ib

```

; TITLE OF INVENTION: CHIA
;
; NUMBER OF SEQUENCES: 11
;
; SOURCE OF SEQUENCE:

```

;  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle

STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester

CITY: ROCHESTER  
STATE: New York  
COUNTRY: USA

COUNTRY: USA  
ZIP: 14603

TITLE OF INVENTION:	COMPOSITIONS AND METHODS FOR THE
TITLE OF INVENTION:	DIAGNOSIS, PREVENTION, AND TREAT
TITLE OF INVENTION:	GROWTH AND PROLIFERATION

; FILING DATE: 16-NOV-1988  
 ; PUBLICATION DATE: 24-MAY-1989  
 ; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 293  
 US-07-821-717B-6

```

Query Match          16.2%   Score 71.5; DB 1; Length 610;
Best Local Similarity 45.0%; Pred. No. 2.7;
Matches 18; Conservative 2; Mismatches 11; Indels

QY 36 PAPLP-----PEDAPNAASLAPTPVSPVLPFPNLTSEPS 69
      ||| ||| ||| ||| ||| ||| ||| |||
DB 366 PTPSPTTSEVPPEPAPNNTLTETPTSPRTTPEP-----TSEPA 402

RESULT 6
US-08-119-262B-6
: Sequence 6, Application US/08119262B
: Patent No. 5492809
: GENERAL INFORMATION:
: APPLICANT: Miller, Jonathan L.
: APPLICANT: Cunningham, David
: APPLICANT: Lyle, Vicki A.
: APPLICANT: Finch, Clara N.
: TITLE OF INVENTION: MUTATIONS RENDERING PLATELET
: TITLE OF INVENTION: GLYCOPROTEIN Ib ALPHA LESS REACTIVE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nixon, Hargrave, Devans & Doyle
: STREET: Clinton Square, P.O. Box 1051
: CITY: Rochester
: STATE: New York
: COUNTRY: USA
: ZIP: 14603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION NUMBER: US/08/119,262B
: APPLICATION DATE: 09-SEP-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/821,717
: FILING DATE: 15-JAN-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Timain, Susan J.
: REGISTRATION NUMBER: 34,103
: REFERENCE/DOCKET NUMBER: 20884/22
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (716) 263-1636
: TELEFAX: (716) 263-1600
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 610 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PUBLICATION INFORMATION:
: AUTHORS: Lopez, Jose A.
: AUTHORS: Chung, Dominic W.
: AUTHORS: Fujikawa, Kazuo
: AUTHORS: Hagen, Frederick S.
: AUTHORS: Papayannopoulou, Thalia
: AUTHORS: Roth, Gerald J.
: TITLE: Cloning of the alpha chain of human platelet
: TITLE: glycoprotein Ib: A transmembrane protein
: TITLE: leucine-rich alpha-2-glycoprotein
: JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
: VOLUME: 84
: PAGES: 5615-5619

```

TITLE: glycoprotein Ib: A transmembrane protein  
 TITLE: leucine-rich alpha-2-glycoprotein  
 JOURNAL: Proc. Natl. Acad. Sci. U.S.A.  
 VOLUME: 84  
 PAGES: 5615-5619

```

Query Match      100.0%; Score 442; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.1e-42;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VRRQLPVEEPNPAKRLLFLLLTIVFCQILMAEEGVPAFLPPEDAPNAASLAPTPVSPYLE 60
      |||||
Db 71 VRRQLPVEEPNPAKRLLFLLLTIVFCQILMAEEGVPAFLPPEDAPNAASLAPTPVSPYLE 130
      |||||

Qy 63 PNLTSEUSDYALDLSTFLQQHPAAF 86
      |||||
Db 131 PNLTSEUSDYALDLSTFLQQHPAAF 156
      |||||

```

RESULT 4  
US-09-356-952-3  
; Sequence 3, Application US/09356952  
; Patent No. 6117663  
; GENERAL INFORMATION:  
; APPLICANT: Borlack-Stodin, Ann  
; APPLICANT: Margarit, S. M.  
; APPLICANT: Bor-Sodit, Dafna  
; APPLICANT: Cole, Philip  
; APPLICANT: Kuriyan, John  
; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: 600-1-228N  
; CURRENT APPLICATION NUMBER: US/09/356,952  
; CURRENT FILING DATE: 1999-07-19  
; EARLIER APPLICATION NUMBER: 60/093,631  
; EARLIER FILING DATE: 1998-07-21  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 1596  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-356-952-3

```
Query Match      18.0%: Score 79.5; DB 3: Length 1596;
Best Local Similarity 28.3%; Pred.No.1.1;
Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps
        6;
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 22:06:25 ; Search time 36.91 seconds  
(without alignments)  
41.840 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156

Perfect score: 442

Sequence: 1 VRRQLPVEEPNPAKRLLELL.....EPSDYALDLSTFLQHPAAF 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/prodata/2/iaa/6\_COMB.pep.\*

4: /cgn2\_6/prodata/2/iaa/PCTUS\_COMB.pep.\*

5: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	442	100.0	156	3	US-08-826-246-10
2	442	100.0	156	3	US-08-944-495-10
3	442	100.0	156	3	US-09-126-640-11
4	79.5	18.0	1596	3	US-09-356-952-3
5	71.5	16.2	610	1	US-07-821-717B-6
6	71.5	16.2	610	1	US-08-119-262B-6
7	71.5	16.2	610	1	US-08-135-929A-11
8	71.5	16.2	610	1	US-08-234-265A-11
9	69.5	15.7	739	3	US-09-035-648-24
10	68.5	15.5	878	1	US-08-237-919-2
11	68.5	15.5	878	4	PCT-US95-05518-2
12	68.5	15.5	1337	3	US-08-854-585-2
13	68.5	15.5	1337	4	PCT-US95-05512-2
14	65.5	14.8	286	1	US-08-382-184-3
15	65.5	14.8	286	2	US-08-641-356-3
16	65.5	14.8	325	1	US-08-382-184-2
17	65.5	14.8	325	2	US-08-641-356-2
18	65.5	14.8	1093	3	US-08-545-860D-55
19	65.5	14.8	1093	4	PCT-US94-04496-55
20	65	14.7	328	2	US-08-651-818A-2
21	64.5	14.6	147	2	US-08-530-165-2
22	64.5	14.6	147	2	US-08-530-165-3
23	64.5	14.6	863	2	US-08-666-271-2
24	63.5	14.4	211	2	US-08-184-292B-2
25	63.5	14.4	211	3	US-08-845-623-2
26	63.5	14.4	211	3	US-08-815-927-2
27	63	14.3	174	2	US-08-683-262B-47
28	62	14.0	652	1	US-08-318-831-8

29	61.5	13.9	166	1	US-07-935-311A-2
30	61.5	13.9	166	1	US-08-368-079-2
31	61.5	13.9	166	4	PCT-US93-07996-2
32	61	13.8	297	2	US-08-580-545B-6
33	61	13.8	297	3	US-08-262-653A-6
34	61	13.8	306	2	US-08-824-707-2
35	61	13.8	486	2	US-08-942-423-3
36	61	13.8	537	3	US-08-920-610-2
37	60.5	13.7	259	3	US-08-469-318-133
38	60.5	13.7	259	3	US-08-468-609A-133
39	60.5	13.7	259	4	PCT-US95-01185-133
40	60.5	13.7	347	2	US-09-004-502-1
41	60.5	13.7	1297	2	US-08-290-731C-4
42	59.5	13.5	259	3	US-08-469-318-131
43	59.5	13.5	259	3	US-08-469-318-132
44	59.5	13.5	259	3	US-08-468-609A-131
45	59.5	13.5	259	3	US-08-468-609A-132

ALIGNMENTS

RESULT 1  
US-08-826-246-10  
; Sequence 10, Application US/08826246  
; Patent No. 6048709  
; GENERAL INFORMATION:  
; APPLICANT: Falb, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/826,246  
; FILING DATE: 28-MAR-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/799,910  
; FILING DATE: 13-FEB-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/011,787  
; FILING DATE: 16-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-078-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)7909090  
; TELEFAX: (212)8699741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 156 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-826-246-10

Search completed: January 30, 2001, 23:09:04  
Job time: 6029 sec

Db 5 KLPLSPSVIRIRILFYLLMLLFCOOLAMIFWRVGLP-----DNSPVASVOITPAQARQQ 58

QY 61 PFNLTSPESDYAL 73  
| | | | | : | : |  
Matches 29; Conservative 13; Mismatches 39; Indels 21; Gaps 5;

Db 59 PVTL-----NDFTL 67

RESULT 12  
S27920  
nuclear antigen EBNA-3A - human herpesvirus 4  
C:Species: human herpesvirus 4, Epstein-Barr virus  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 21-Jul-2000  
C:Accession: S27920; G49253; H49253; F49253  
R:Sample, J.; Young, L.; Martin, B.; Chatman, T.; Kieff, E.; Rickinson, A.; Kieff, E.  
submitted to the EMBL Data Library, July 1990  
A:Reference number: S27920  
A:Accession: S27920  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-925 <SAM>  
A:Cross-references: EMBL:M34440; NID:g330407; PIDN:AAA45893.1; PID:g330408  
R:Apolloni, A.; Moss, D.; Stumm, R.; Burrows, S.; Suhrbier, A.; Misko, I.; Schmidt, C.;  
Eur. J. Immunol. 22, 183-189, 1992  
A:Title: Sequence variation of cytotoxic T cell epitopes in different isolates of Epstein-Barr virus  
A:Reference number: A49034; MUID:92111623  
A:Accession: G49253  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 314-332 <APO>  
A:Cross-references: GB:S79235; NID:g242901; PIDN:AB20990.1; PID:g242902  
A:Experimental source: human B-type strain QIMR-JSM6  
A:Note: sequence extracted from NCBI backbone (NCBIN:79235, NCBI:P:79252)  
A:Accession: H49253  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 314-338 <AP2>  
A:Cross-references: GB:S79234; NID:g242899; PIDN:AB20989.1; PID:g242900  
A:Experimental source: human B-type strain AG876JS; human B-type strain L4MR  
A:Note: sequence extracted from NCBI backbone (NCBIN:79237, NCBI:P:79254, NCBIN:79234, NCBI:P:79255)  
A:Genetics:  
A:Introns: 112/3  
C:Superfamily: human herpesvirus 4 nuclear antigen EBNA-3A

Query Match 16.6%; Score 73.5; DB 2; Length 925;  
Best Local Similarity 28.4%; Pred. No. 17;  
Matches 29; Conservative 13; Mismatches 39; Indels 21; Gaps 5;

QY 1 VRRQLPVEEPNPAKRLFLLLTIVFCOI--LMAEEGVAPALP-----PEDA 44  
| | | | | : | : | | | | |  
Db 606 VSPQPMERPLEPQOMF--PGSPFSQVADVARESGVPAMQPOYFDLPLTQPIISOGAPAA 663

QY 45 PNAASLAPTPVSPVLEP--ENL-TSEPSDYALDLSTFLOQHP 83  
| | | | | : | : | | | | |  
Db 564 PLRASMGPVPVPVTPQOYFDIPLTEPINOGASAAHFLPOOP 705

RESULT 13  
T30351  
mucin-like protein - Lymantria dispar nuclear polyhedrosis virus  
C:Species: Lymantria dispar nuclear polyhedrosis virus, LdMNPV  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T30351  
R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rohrer, J.L.;  
Virology 253, 17-34, 1999  
A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria dispar  
A:Reference number: 220836; MUID:99124785  
A:Accession: T30351  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1029 <KUZ>  
A:Cross-references: EMBL:AF081810; PIDN:AAC70189.1

Query Match 16.5%; Score 73; DB 2; Length 1029;  
Best Local Similarity 38.2%; Pred. No. 22;  
Matches 29; Conservative 3; Mismatches 30; Indels 14; Gaps 4;

QY 6 PVEEPNPAKRLFLLLTIVFCOILMAEEGVAPALP---PEDAPNAASL-----APTVPVSP 57  
| | | | | : | : | | | | |  
Db 644 PVPEPS-APVDYFTLSAEF-----APEPAPEPAPEPAPEPTSRFTSEPALGPVEP 697

QY 58 VLPEFNLTSEPSDYAL 73  
| | | | | : | : |  
Db 698 ALEPVEPALEPVEPAL 713

RESULT 14  
S28013  
outC protein - Erwinia chrysanthemi  
C:Species: Erwinia chrysanthemi  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 21-Jul-2000  
C:Accession: S28013; S23885  
R:Condemine, G.; Dorel, C.; Hugouvieux-Cotte-Pattat, N.; Robert-Houdouy, J.;  
Mol. Microbiol. 6, 3199-3211, 1992  
A:Title: Some of the out genes involved in the secretion of pectate lyases in Erwinia chrysanthemi  
A:Reference number: S28011; MUID:93086427  
A:Accession: S28013  
A:Molecule type: DNA  
A:Residues: 1-272 <CON>  
A:Cross-references: EMBL:X65265; NID:g3152953; PIDN:CAA46369.1; PID:g42201  
C:Genetics:  
A:Gene: outC  
C:Keywords: transmembrane protein

Query Match 16.4%; Score 72.5; DB 2; Length 272;  
Best Local Similarity 30.1%; Pred. No. 5.6;  
Matches 22; Conservative 16; Mismatches 22; Indels 13; Gaps 4;

QY 4 QLPVEEPNPAKRLFLLLTIVFCO---ILMAEEGVAPALPPEPADAPNAASLAPTPVSPVLE 60  
| | | | | : | : | | | | |  
Db 5 KLPLSPSVIRIRILFYLLMLLFCOOLAMIFWRIGL-----PDNAP-VSSVOITPAQARQQ 58

QY 61 PFNLTSPESDYAL 73  
| | | | | : | : |  
Db 59 PVTL-----NDFTL 67

RESULT 15  
NBHUIA  
platelet glycoprotein Ib alpha chain precursor - human  
N:Alternate names: membrane glycoprotein Ib alpha chain  
N:Contains: glycosylated  
C:Species: Homo sapiens (man)  
C:Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 22-Jun-1999  
C:Accession: A94174; A60435; A94173; S16945; I55355; A27075; A27102  
R:Lopez, J.A.; Chung, D.W.; Fujikawa, K.; Hagen, F.S.; Papayannopoulou, T.; Roth,  
Proc. Natl. Acad. Sci. U.S.A. 84, 5615-5619, 1987  
A:Title: Cloning of the alpha-chain of human platelet glycoprotein Ib: a transmembrane protein  
A:Reference number: A94174; MUID:87289655  
A:Accession: A94174  
A:Molecule type: mRNA  
A:Residues: 1-626 <LOP>  
A:Cross-references: GB:J02940; NID:g183499; PIDN:AAA52595.1; PID:g306793  
R:Wicki, A.N.; Walz, A.; Gerber-Huber, S.N.; Wenger, R.H.; Vornhagen, R.; Clemetson,  
Thromb. Haemost. 61, 448-453, 1989  
A:Title: Isolation and characterization of human blood platelet mRNA and construction of a cDNA library  
A:Reference number: A60435; MUID:90020160  
A:Accession: A60435  
A:Molecule type: mRNA  
A:Residues: 207-467 <WIC>  
R:Titani, K.; Takio, K.; Handa, M.; Ruggeri, Z.M.  
Proc. Natl. Acad. Sci. U.S.A. 84, 5610-5614, 1987  
A:Title: Amino acid sequence of the von Willebrand factor-binding domain of platelet glycoprotein Ib



QY 61 PF-NLTSEP 68  
| | | | |  
Db 4657 PLVELPTP 4665

RESULT 7  
A81141  
acyl CoA thioester hydrolase family protein NMB0925 [imported] - Neisseria meningitidis  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 31-Mar-2000  
C:Accession: A81141  
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
Zi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.;  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755  
A:Accession: A81141  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-148 <TET>  
A:Cross-references: GB:AB002444; GB:AB002098; NID:g7226162; PIDN:AAF41333.1; PID:g722616  
A:Experimental source: serogroup B, strain NMD58  
C:Genetics:  
A:Gene: NMB0925

Query Match 17.2%; Score 76; DB 2; Length 148;  
Best Local Similarity 40.5%; Pred. No. 1.3;  
Matches 15; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY 18 FULLTIVFCQILMAEGVPAPLPEDAPNAASLAPTP 54  
| | | | |  
Db 112 YLVTEAVFTYVAIDAEGRPRIPKGNPKLAGLLPTP 148  
| | | | |

RESULT 8  
T18344  
P-glycoprotein E - Leishmania tropica  
C:Species: Leishmania tropica  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T18344  
R:Laudente, E.; Castansys, S.; Gamarro, F.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z18880  
A:Accession: T18344  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1677 <LAF>  
A:Cross-references: EMBL:U55381; NID:g1916605; PID:g1916606; PIDN:AAB51191.1  
C:Superfamily: human multidrug resistance protein cMOAT2; ATP-binding cassette homology

Query Match 17.2%; Score 76; DB 2; Length 1677;  
Best Local Similarity 33.9%; Pred. No. 19;  
Matches 20; Conservative 6; Mismatches 15; Indels 18; Gaps 2;

QY 37 APLP-----PEDAPNAASLAPTPVSPVLEPFN-----LTSEPSDYALDLST 77  
| | | | |  
Db 698 APLPEAEFLPRQPSSSAARPRAGPVTEPLNKGSHTGCHDAAASSEPLSSSAQKST 756  
| | | | |

RESULT 9  
T00247  
zinc finger protein wiz - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 05-Nov-1999  
C:Accession: T00247  
R:Matsumoto, K.; Ishii, N.; Yoshida, S.; Shiosaka, S.; Wanaka, A.; Tohyama, M.  
submitted to the EMBL Data Library, March 1998  
A:Description: Molecular cloning and distinct developmental expression pattern of splice

A:Reference number: Z14130  
A:Accession: T00247  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-955 <MAT>  
A:Cross-references: EMBL:AB012266; NID:d1227741; PIDN:BAA32791.1; PID:d1033757  
A:Experimental source: brain  
C:Genetics:  
A:Gene: wiz

Query Match 16.7%; Score 74; DB 2; Length 955;  
Best Local Similarity 41.9%; Pred. No. 16;  
Matches 18; Conservative 2; Mismatches 15; Indels 8; Gaps 1;

QY 34 GVPAPLPEDAPNAASLAPTPVSPVLE-----PFNLITSEP 68  
| | | | |  
Db 300 GSPTPKNPEDKSPQLSLSPRTSPRAOCPQSEDEGLNLTSGP 342  
| | | | |

RESULT 10  
T00248  
zinc finger protein wizL - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 05-Nov-1999  
C:Accession: T00248  
R:Matsumoto, K.; Ishii, N.; Yoshida, S.; Shiosaka, S.; Wanaka, A.; Tohyama, M.  
submitted to the EMBL Data Library, March 1998  
A:Description: Molecular cloning and distinct developmental expression pattern o  
A:Reference number: Z14130  
A:Accession: T00248  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1561 <MAT>  
A:Cross-references: EMBL:AB012265; NID:d1227740; PIDN:BAA32790.1; PID:d1033756  
A:Experimental source: brain  
C:Genetics:  
A:Gene: wiz

Query Match 16.7%; Score 74; DB 2; Length 1561;  
Best Local Similarity 41.9%; Pred. No. 27;  
Matches 18; Conservative 2; Mismatches 15; Indels 8; Gaps 1;

QY 34 GVPAPLPEDAPNAASLAPTPVSPVLE-----PFNLITSEP 68  
| | | | |  
Db 906 GSPTPKNPEDKSPQLSLSPRTSPRAOCPQSEDEGLNLTSGP 948  
| | | | |

RESULT 11  
A47021  
pectic enzyme secretion protein OutC - Erwinia chrysanthemi  
C:Species: Erwinia chrysanthemi  
C:Date: 24-Feb-1994 #sequence\_revision 18-Nov-1994 #text\_change 16-Feb-1997  
C:Accession: A47021  
R:Lindeberg, M.; Collmer, A.  
J. Bacteriol. 174, 7385-7397, 1992  
A:Title: Analysis of eight out genes in a cluster required for pectic enzyme sec  
A:Reference number: A47021; MUID:93054355  
A:Accession: A47021  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-272 <LIN>  
A:Note: sequence extracted from NCBI backbone (NCBIP:118270)  
C:Keywords: transmembrane protein

Query Match 16.6%; Score 73.5; DB 2; Length 272;  
Best Local Similarity 30.1%; Pred. No. 4.5;  
Matches 22; Conservative 13; Mismatches 25; Indels 13; Gaps 3;

QY 4 QLPVEEPNPAKRLFLITIVFCQ---ILMAEEGVAPAPLPEDAPNAASLAPTPVSPVLE 60  
| | | | |

	Matches	58:	Conservative	5:	Mismatches	14:	Indels	8:	Gaps	2:
Qy	1	VRRQLVPEEPNPARKLFLLLTI	VFCQLMAEEGVAPLP	PEDAPNAASLAPTVS----	56					
Dd	71	KKRLITTEPNIARVFLLFAII	FICQLMAEEGVSQPLA	EDATSA--VTPEPISAPIT	128					
Qy	57	--PVLEPFNLTSEPSYALDLSTFL	79							
Dd	129	APPVLEPLNTTSSDYALDKAF	153							

```

RESULT      3
T33565
Hypothetical protein R160.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T33565
R: Nelson, J.; Gattung, S.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of C. elegans cosmid R160.
A:Reference number: 221370
A:Accession: T33565
A:status: preliminary; translated from GB/EMBL/DBJ
A:molecule type: DNA
A:Residues: 1-102 <NLE>
A:Cross-references: EMBL:AF099001; PIDN:AAC68734.1; GSPDB:GNO0028; CESP:R160
A:Experimental source: strain Bristol N2; clone R160
C:Genetics:
A:Gene: CESP:R160.5
A:Map position: X
A:Introns: 25/2

```

```

Query Match      18.4%; Score 81.5; DB 2; Length 102;
Best Local Similarity 24.7%; Pred. No. 0.26;
Matches 19; Conservative 19; Mismatches 28; Indels 11; Gaps 3;

QY    11 NPAKRLLELL-----LTIVFCQLMAEEGVAPALPPEDAPNAASLAPTVPSPVLEPN-- 63
      ||| :||: :||: :||: |||||: :||: :||: :||:
Db     7 NPATQAFVVILCALAVCICMLIERADGFCCPLPPEELRIVTIP-PRTPIAAPVNRC 65

QY    64 ---LTSEPSDYALDLST 77

Db     66 QVVIKIDPTESDVDMTS 82

```

RESULT 4  
A41216  
guanine nucleotide exchange activator son-of-sevenless - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C:Date: 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 24-Sep-1998  
C:Accession: A41216  
R:Simon, M.A.; Bowtell, D.D.L.; Dodson, G.S.; Laverty, T.R.; Rubin, G.M.  
Cell 67, 701-716, 1991  
A:Title: Rasl and a putative guanine nucleotide exchange factor perform crucial steps in  
A:Reference number: A41216; MUID:92034991  
A:Accession: A41216  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1596 <SIM>  
A:Cross-references: GB:M77501; NID:q158470; PID:q158471  
C:Genetics:  
A:Gene: FlyBase:Sos  
A:Cross-references: FlyBase:FBgn0001965  
C:Superfamily: CDC35-type guanine nucleotide exchange activator homology; plectstrin ref  
F:479-586/Domain: plectstrin repeat homology <PLK>  
F:825-1066/Domain: CDC35-type guanine nucleotide exchange activator homology <SOS>

Query Match	18.0%	Score 79.5;	DB 2;	Length 1596;
Best Local Similarity	28.3%	Pred. No. 8.2;		
Matches	30;	Conservative	14;	Mismatches 37;
			Indels	25;
			Gaps	6;

```

QY 1 VRRQLPVE-----EPNPAKRLFLFLLLTIVFCQILMAEEGVPAIPDPEDA
DB 1423 IRRNSAIKRAAATOPNOAGAPISITLTVTVSQAVATDEPLPLPISP-----
QY 54 ---PVSPVLPEPF---NITSEP-----SDYALDLSTFLQQ-----HPAAF 86
DB 1478 TSPLPKAMPSPNPISHPHVESTSSSYAHQLMRQOQOQOQTHIPAIV 1523

```

```

RESULT      5
T03455
ALR protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03455
R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli,
Oncogene 15, 549-560, 1997
A:Title: Structure and expression pattern of human ALR, a novel gene with stro
A:Reference number: Z14954; MUID:97388474
A:Accession: T03455
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-4957 <PRA>
A:Cross-references: EMBL:AF010404; NID:g2358286; PID:g2358287
C:Genetics:
A:Gene: ALR
A:Map position: 12
C:Superfamily: human ALR protein
C:Keywords: alternative splicing

```

Query Match 17.6%; Score 78; dB 2; Length 4957;  
Best Local Similarity 30.4%; Pred. No. 39;  
Matches 21; Conservative 12; Mismatches 28. Indels

```

QY      1 VRQLPVEEPNPAKRLFLLLTVFCQILMAEGVPAFLPPEDAPNAASLAPTEVSPVLE 60
      : : : : | | | | : : : : | | | | : : : : | | | | : : : : | | | |
Db 4299 LKQESAPPEPTQHRITYNVSNDLVRQL-----SAPPEEPSPPLAPSAPPTE 4351
      : : : : | | | | : : : : | | | | : : : : | | | | : : : : | | | |
QY      61 PF-NLTSEP 68
      : : : : | | | | : : : : | | | | : : : : | | | | : : : : | | | |
Db 4352 PLVELPTEP 4360
      : : : : | | | | : : : : | | | | : : : : | | | | : : : : | | | |

```

RESULT 6

T03454

ALR protein - human

C:Species: Homo sapiens (man)

C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000

C:Accession: T03454

C:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallepalli, V.

C:Title: Structure and expression pattern of human ALR, a novel gene with strong

A:Reference number: Z14954; MUID:97388474

A:Accession: T03454

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-5262 <PRA>

A:Cross-references: EMBL:AF010403; NID:g2358284; PID:g2358285

C:Genetics:

A:Gene: ALR

A:Map position: 12

C:Superfamily: human ALR protein

C:Keywords: alternative splicing

```

Query Match      17.6%  Score 78;  DB 2;  Length 5362;
Best Local Similarity 30.4%  Pred. No. 42;
Matches 21;  Conservative 12;  Mismatches 28;  Indels 8;  Gaps 2;
QY 1 VRROLPEVERNPARKLFLLLTIYFCQILMAEEGVVAPLPPEDPAPNAASLAPTPVSPVLE 60
      :::: ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4604 LKQESAPPEPTQHRYYNVNLDVROL-----SAPPEEPSPPPSLAPSPASPTPE 4656

```

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 23:09:03 ; Search time 21.17 Seconds  
(without alignments)  
275.836 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156  
Perfect score: 442  
Sequence: 1 VRRQLPVEEPNPAKRLLELL.....EPFDYALDLSTFLOQHPAAF 86

Scoring table:  
BLOSUMP2  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	442	100.0	156	2 JC5537	differentiation-de
2	273	61.8	153	2 S33363	gly96 protein - mo
3	81.5	18.4	102	2 T33565	hypothetical prote
4	79.5	18.0	1596	2 A41216	guanine nucleotide
5	78	17.6	4957	2 T03455	ALR protein - huma
6	78	17.6	5262	2 T03454	ALR protein - huma
7	76	17.2	148	2 A81141	acyl CoA thioester
8	76	17.2	1677	2 T18344	P-glycoprotein E -
9	74	16.7	955	2 T00247	zinc finger protei
10	74	16.7	1561	2 T00248	zinc finger protei
11	73.5	16.6	272	2 A47021	pectic enzyme secr
12	73.5	16.6	925	2 S27920	nuclear antigen EB
13	73	16.3	1029	2 T30351	mucin-like protein
14	72.5	16.4	272	2 S28013	outc protein - Erw
15	71.5	16.2	626	1 NBH7A	platelet glycoprot
16	71	16.1	359	2 T33756	hypothetical prote
17	70.5	16.0	235	2 D83477	hypothetical prote
18	70.5	16.0	605	1 Q0BE29	hypothetical prote
19	70	15.8	148	2 E81878	BRLFI protein - hu
20	70	15.8	1233	2 T37045	probable acyl-CoA
21	69.5	15.7	1887	2 S61703	nitrate reductase
22	69	15.6	418	2 T19800	fatty-acid synthas
23	69	15.6	1402	2 T46707	hypothetical prote
24	69	15.6	2187	2 T30826	translation initia
25	68.5	15.5	181	2 T08793	nascent polypeptid
26	68.5	15.5	882	1 TJHUCE	hypothetical prote
27	68.5	15.5	1072	2 T50949	cadherin 1 precurs
28	68.5	15.5	1337	1 I38670	verprolin related
29	68.5	15.5	1844	2 S01956	protein-tyrosine-p
					hypothetical prote

30	68.5	15.5	2649	2 T51023	hypothetical prote
31	68	15.4	241	2 T22216	hypothetical prote
32	68	15.4	267	2 A49068	cranosynotosis-as
33	68	15.4	1396	2 A44453	translation initia
34	67	15.2	283	2 S13383	hydroxyproline-ric
35	67	15.2	586	2 S19381	hypothetical prote
36	67	15.2	3566	1 A40701	tenascin-x precurs
37	66.5	15.0	169	2 A72466	hypothetical prote
38	66.5	15.0	3149	1 Q0BE8	BPLFI protein - hu
39	66	14.9	101	2 T46506	hypothetical prote
40	66	14.9	187	2 C75558	acetyl-CoA carboxy
41	66	14.9	621	2 JC5164	acetylacate synth
42	66	14.9	5149	2 F83345	probable non-ribos
43	65.5	14.8	325	2 D70666	exBB protein - pse
44	65.5	14.8	329	2 S28442	hypothetical prote
45	65.5	14.8	424	2 T43468	

ALIGNMENTS

RESULT 1  
JC5537  
differentiation-dependent protein DIF-2 - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Sep-1997 #sequence\_revision 05-Sep-1997 #text\_change 07-May-1999  
C:Accession: JC5537  
R:Pietzsch, A.; Buechler, C.; Aslanidis, C.; Schmitz, G.  
Biochem. Biophys. Res. Commun. 235, 4-9, 1997  
A:Title: Identification and characterization of a novel monocyte/macrophage diff  
A:Reference number: JC5537; MUID:97339426  
A:Accession: JC5537  
A:Molecule type: mRNA  
A:Residues: 1-156 <PIE>  
A:Experimental source: monocyte  
A>Note: the authors translated the codon CCG for residue 106 as Arg

Query Match 100.0%; Score 442; DB 2; Length 156;  
Best Local Similarity 100.0%; Pred. No. 6.2e-36;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VRRQLPVEEPNPAKRLLELLTIVFCQILMAEGVPAPLPPEADPNAAASLAPTVPSPVLE 60  
|||||  
Db 71 VRRQLPVEEPNPAKRLLELLTIVFCQILMAEGVPAPLPPEADPNAAASLAPTVPSPVLE 130  
Qy 61 PFNLTSEPSDYALDLSTFLOQHPAAF 86  
|||||  
Db 131 PFNLTSEPSDYALDLSTFLOQHPAAF 156

RESULT 2  
S33363  
gly96 protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Feb-1997  
C:Accession: S33363  
R:Charles, C.H.; Yoon, J.K.; Sinske, J.S.; Lau, L.F.  
Oncogene 8, 797-801, 1993  
A:Title: Genomic structure, cDNA sequence, and expression of gly96, a growth fac  
A:Reference number: S33363; MUID:93173526  
A:Accession: S33363  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-153 <CHA>  
A:Cross-references: EMBL:X67644  
C:Genetics: 70/3  
C:Keywords: transmembrane protein

Query Match 61.8%; Score 273; DB 2; Length 153;  
Best Local Similarity 68.2%; Pred. No. 1.3e-19;

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 21:28:30 ; Search time 49.1 seconds  
(without alignments)  
846.830 Million cell updates/sec

Title: US-08-799-910-9\_COPY\_211\_468

Perfect score: 258  
Sequence: 1 GTCCGGCCCGCCAGTCCAGT.....AGCAACACCCGGCCGCTTC 258

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/prodata/2/ina/6\_COMB.seq.\*  
4: /cgn2\_6/prodata/2/ina/PCUS\_COMB.seq.\*  
5: /cgn2\_6/prodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	258	100.0	1228	3	US-08-826-246-9 Sequence 9, Appli
2	258	100.0	1228	3	US-08-944-495-9 Sequence 9, Appli
3	258	100.0	1228	3	US-09-126-640-5 Sequence 5, Appli
4	43	16.7	7218	1	US-08-232-463-14 Sequence 14, Appli
5	39.4	15.3	2847	1	US-08-087-007-2 Sequence 2, Appli
6	39.4	15.3	2847	3	US-08-483-433-2 Sequence 2, Appli
7	39.4	15.3	2847	4	PCT-US92-05920-2 Sequence 2, Appli
8	38.6	15.0	6354	3	US-09-058-389A-5 Sequence 5, Appli
9	37.6	14.6	170	3	US-09-058-389A-13 Sequence 13, Appli
10	34.8	13.5	2953	2	US-08-859-201-1 Sequence 1, Appli
11	33.6	13.0	1120	1	US-08-188-582-8 Sequence 8, Appli
12	33.6	13.0	1120	1	US-08-546-715-8 Sequence 8, Appli
13	32	12.4	3765	3	US-07-705-490-1 Sequence 1, Appli
14	32	12.4	4362	2	US-08-455-073A-1 Sequence 1, Appli
15	31.8	12.3	599	1	US-08-584-226-3 Sequence 3, Appli
16	31.6	12.2	2074	2	US-09-018-576-2 Sequence 2, Appli
17	31.6	12.2	2074	2	US-09-018-576-11 Sequence 11, Appli
18	31.6	12.2	2074	3	US-09-248-137-2 Sequence 2, Appli
19	31.6	12.2	2074	3	US-09-248-137-11 Sequence 11, Appli
20	31.4	12.2	1335	4	PCT-US91-06532-1 Sequence 1, Appli
21	31.4	12.2	38506	3	US-09-320-878-19 Sequence 19, Appli
22	30.8	11.9	797	1	US-08-332-467B-2 Sequence 2, Appli
23	30.8	11.9	797	1	US-08-681-811-2 Sequence 2, Appli
24	30.8	11.9	797	4	PCT-US93-12507-2 Sequence 2, Appli
25	30.8	11.9	2680	3	US-09-063-035-1 Sequence 1, Appli
26	30.6	11.9	7791	3	US-08-949-386-23 Sequence 23, Appli
27	30.6	11.9	7791	3	US-08-450-562-23 Sequence 23, Appli
28	30.6	11.9	7808	3	US-08-949-386-22 Sequence 22, Appli

C	29	30.6	11.9	7808	3	US-08-450-562-22 Sequence 22, Appli
	30	30.2	11.7	22481	4	PCT-US95-07201-43 Sequence 43, Appli
	31	30	11.6	5688	5	5248670-4 Patent No. 5248670
	32	29.8	11.6	536	1	US-08-329-704-1 Sequence 1, Appli
	33	29.8	11.6	536	2	US-08-472-604-1 Sequence 1, Appli
	34	29.8	11.6	536	2	US-08-486-117-1 Sequence 1, Appli
	35	29.8	11.6	538	2	US-08-865-337A-3 Sequence 3, Appli
	36	29.8	11.6	2196	2	US-08-865-337A-2 Sequence 2, Appli
	37	29.8	11.6	8299	1	US-08-462-014-2 Sequence 2, Appli
	38	29.8	11.6	8299	3	US-08-923-137-3 Sequence 3, Appli
	39	29.6	11.5	693	1	US-08-168-091A-3 Sequence 3, Appli
C	40	29.6	11.5	1219	4	PCT-US93-06251-11 Sequence 11, Appli
	41	29.6	11.5	1313	1	US-08-176-427B-7 Sequence 7, Appli
	42	29.6	11.5	1313	2	US-08-356-060A-4 Sequence 4, Appli
	43	29.6	11.5	1313	3	US-08-460-900C-4 Sequence 4, Appli
	44	29.6	11.5	1489	3	US-08-836-582-1 Sequence 1, Appli
	45	29.6	11.5	2485	1	US-08-424-424B-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-826-246-9  
: Sequence 9, Application US/08826246  
: Patent No. 6048709  
: GENERAL INFORMATION:  
: APPLICANT: Falb, Dean  
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
: TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
: TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
: NUMBER OF SEQUENCES: 44  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: PERNIE & EDMONDS LLP  
: STREET: 1155 Avenue of the Americas  
: CITY: New York  
: STATE: NY  
: COUNTRY: USA  
: ZIP: 10036-2711  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette  
: COMPUTER: IBM Compatible  
: OPERATING SYSTEM: DOS  
: SOFTWARE: FastSeq Version 2.0  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/826,246  
: FILING DATE: 28-MAR-1997  
: CLASSIFICATION: 800  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 08/799,910  
: FILING DATE: 13-FEB-1997  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 60/011,787  
: FILING DATE: 16-FEB-1996  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Coruzzi, Laura A  
: REGISTRATION NUMBER: 30,742  
: REFERENCE/DOCKET NUMBER: 7853-078-999  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (212)7909090  
: TELEFAX: (212)8699741  
: TELEX: 66141 PENNIE  
: INFORMATION FOR SEQ ID NO: 9:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1228 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: both  
: TOPOLOGY: linear  
: MOLECULE TYPE: CDNA  
: FEATURE:  
: NAME/KEY: Coding Sequence  
: LOCATION: 1...468  
: OTHER INFORMATION:



309 caacacccggccgccttc 526

309 caacacccggccgccttc 526

Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTCCGGCGCCAGCTGCCAGTCGAGGAACCGAACCCAGCAAAAGCCCTTCCTTCCTG 60
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Nb 269 gtccggcgccagctgccagtcgaggaccgcaaccagcccaaaaggcttcttctgctg 328
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCGCGCCCTTCGCT 120
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 329 ctcaccatcgcttctgcccagatcctgatggtgaagagggtgtgcggcgccctgct 388
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 121 CCAGAGACGCCCTTAACGCCGCATCCTGGCGCCACCCCTGTCTCCCGCTCTCCAG 180
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 389 ccagaggagcccttaacgccgcacccctggtgcgcccacccctgtgtcccccglcctcag 448
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCTCCAG 240
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 449 ccccttaatctgacttcggagccctcgactacgtctgtggacctcagcacttctctccag 508
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 241 CAACACCGCGCGCCTTC 258
  |||||||||||||||||||
Db 509 caacacccggcgcccttc 526
  |||||||||||||||||||
```

```
RESULT 12
US-09-721-588-4548
; Sequence 4548, Application US/09721588
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Villevial, Jean-Luc
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600-2046-001
; CURRENT APPLICATION NUMBER: US/09/721,588
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: 60/167,381
; PRIOR FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 5410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4548
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-721-588-4548
```

Query Match 100.0%; Score 258; DB 24; Length 673;  
Best Local Similarity 100.0%; Pred. No. 7.7e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTCCGGCGCCAGCTGCCAGTCGAGGAACCGAACCCAGCAAAAGCCCTTCCTTCCTG 60
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 269 gtccggcgccagctgccagtcgaggaccgcaaccagcccaaaaggcttcttctgctg 328
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCGCGCCCTTCGCT 120
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 329 ctcaccatcgcttctgcccagatcctgatggtgaagagggtgtgcggcgccctgct 388
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 121 CCAGAGACGCCCTTAACGCCGCATCCTGGCGCCACCCCTGTCTCCCGCTCTCCAG 180
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 389 ccagaggagcccttaacgccgcacccctggtgcgcccacccctgtgtcccccglcctcag 448
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCTCCAG 240
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 449 ccccttaatctgacttcggagccctcgactacgtctgtggacctcagcacttctctccag 508
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 241 CAACACCGCGCGCCTTC 258
  |||||||||||||||||||
Db 509 caacacccggcgcccttc 526
  |||||||||||||||||||
```

RESULT 13  
US-09-726-787-2982

```
RESULT 10
US-09-710-286-3772
; Sequence 3772, Application US/09710286
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600-2005-001
; CURRENT APPLICATION NUMBER: US/09/710,286
; CURRENT FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: 60/164,255
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4115
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3772
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-710-286-3772
```

```
Query Match 100.0%; Score 258; DB 24; Length 673;  
Best Local Similarity 100.0%; Pred. No. 7.7e-52;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GTCCGGCGCCAGCTGCCAGTCGAGGAACCGAACCCAGCAAAAGCCCTTCCTTCCTG 60
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 269 gtccggcgccagctgccagtcgaggaccgcaaccagcccaaaaggcttcttctgctg 328
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGTGTGCGCGCCCTTCGCT 120
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 329 ctcaccatcgcttctgcccagatcctgatggtgaagagggtgtgcggcgccctgct 388
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 121 CCAGAGACGCCCTTAACGCCGCATCCTGGCGCCACCCCTGTCTCCCGCTCTCCAG 180
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 389 ccagaggagcccttaacgccgcacccctggtgcgcccacccctgtgtcccccglcctcag 448
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACATTTCTCTCCAG 240
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 449 ccccttaatctgacttcggagccctcgactacgtctgtggacctcagcacttctctccag 508
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 241 CAACACCGCGCGCCTTC 258
  |||||||||||||||||||
Db 509 caacacccggcgcccttc 526
  |||||||||||||||||||
```

```
RESULT 11
US-09-716-473-2500
; Sequence 2500, Application US/09716473
; GENERAL INFORMATION:
; APPLICANT: Williamson, Mark
; APPLICANT: Shvjan, Andrew W.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600-2043-001
; CURRENT APPLICATION NUMBER: US/09/716,473
; CURRENT FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/166,502
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 2933
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2500
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-716-473-2500
```

Query Match 100.0%; Score 258; DB 24; Length 673;  
Best Local Similarity 100.0%; Pred. No. 7.7e-52;

```
; Sequence 2982, Application US/09726787
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2010-001
; CURRENT APPLICATION NUMBER: US/09/726,787
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/168,132
; PRIOR FILING DATE: 1999-11-30
; NUMBER OF SEQ ID NOS: 3241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2982
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-726-787-2982
```

```
Query Match 100.0%; Score 258; DB 55; Length 673;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GTCCGGCGCCAGTGCAGTCGAGGAACCGAACCCACCCAAAGGCTTCCTTTCTGCTG 60
   |||||||
Db 269 gtccggcgccagctgcagtcgaggaacccgacccgacccgacccgacccgacccg 328

QY 61 CTCACCATCTCTTCTCCAGATCCCTGATGGCTGAAGAGGGTGTGCGGGGCCCTGCCT 120
   |||||||
Db 329 ctccaccatctcttctgacagatccctgatggtgaagaggtgtgctgctgctgctgct 388

QY 121 CCAGAGACGCCCTTAACGCCGATCCCTGGCGCCACCCCTGTGCTCCCGCTCCGAG 180
   |||||||
Db 389 ccagagagacccctaacgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 448

QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTCGACCTCAGCACTTTCTCTCCAG 240
   |||||||
Db 449 ccctttaatctgacttcggagccctcgactacgcttgagacccgacccgacccgaccc 508

QY 241 CAACACCGCGCGCCTTC 258
   |||||||
Db 509 caacacccgcccgccttc 526
```

```
RESULT 14
US-09-699-998-1778
; Sequence 1778, Application US/09699998
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gearing, David P.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2008-001
; CURRENT APPLICATION NUMBER: US/09/699,998
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/162,362
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 10905
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1778
; LENGTH: 706
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(706)
; OTHER INFORMATION: n = A,T,C or G
US-09-699-998-1778
```

```
Query Match 100.0%; Score 258; DB 23; Length 706;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;
```

```
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCCGGCGCCAGTGCAGTCGAGGAACCGAACCCACCCAAAGGCTTCCTTTCTGCTG 60
   |||||||
Db 195 gtccggcgccagctgcagtcgaggaacccgacccgacccgacccgacccgacccg 254

QY 61 CTCACCATCTCTTCTCCAGATCCCTGATGGCTGAAGAGGGTGTGCGGGGCCCTGCCT 120
   |||||||
Db 255 ctccaccatctcttctgacagatccctgatggtgaagaggtgtgctgctgctgctgct 314

QY 121 CCAGAGACGCCCTTAACGCCGATCCCTGGCGCCACCCCTGTGCTCCCGCTCCGAG 180
   |||||||
Db 315 ccagagagacccctaacgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 374

QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTCGACCTCAGCACTTTCTCTCCAG 240
   |||||||
Db 375 ccctttaatctgacttcggagccctcgactacgcttgagacccgacccgacccgaccc 434

QY 241 CAACACCGCGCGCCTTC 258
   |||||||
Db 435 caacacccgcccgccttc 452
```

```
RESULT 15
US-09-699-998-10469
; Sequence 10469, Application US/09699998
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gearing, David P.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2008-001
; CURRENT APPLICATION NUMBER: US/09/699,998
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/162,362
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 10905
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10469
; LENGTH: 803
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-699-998-10469
```

```
Query Match 100.0%; Score 258; DB 23; Length 803;
Best Local Similarity 100.0%; Pred. No. 7.8e-52;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCCGGCGCCAGTGCAGTCGAGGAACCGAACCCACCCAAAGGCTTCCTTTCTGCTG 60
   |||||||
Db 184 gtccggcgccagctgcagtcgaggaacccgacccgacccgacccgacccgacccg 243

QY 61 CTCACCATCTCTTCTCCAGATCCCTGATGGCTGAAGAGGGTGTGCGGGGCCCTGCCT 120
   |||||||
Db 244 ctccaccatctcttctgacagatccctgatggtgaagaggtgtgctgctgctgctgct 303

QY 121 CCAGAGACGCCCTTAACGCCGATCCCTGGCGCCACCCCTGTGCTCCCGCTCCGAG 180
   |||||||
Db 304 ccagagagacccctaacgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 363

QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTCGACCTCAGCACTTTCTCTCCAG 240
   |||||||
Db 364 ccctttaatctgacttcggagccctcgactacgcttcgagacccgacccgacccgaccc 423

QY 241 CAACACCGCGCGCCTTC 258
   |||||||
Db 424 caacacccgcccgccttc 441
```

```
Search completed: January 30, 2001, 22:04:50
Job time: 13154 sec
```



ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5876949rls  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: PA USA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455.073A  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca, Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: UPN-2201  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4362 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 220..2118  
US-08-455-073A-1

Query Match 12.4%; Score 32; DB 2; Length 4362;

Best Local Similarity 49.5%; Pred. No. 4.2;  
Matches 109; Conservative 0; Mismatches 110; Indels 1; Gaps 1;

QY 28 CGCAACCCAGCCAAAGGCTTCTTCTGCTGCTCACCATCGTCTTCTGCGCAGATCTCG 87  
DB 246 CGGCACCTCCACACACAGCTCTCCATCTTCTTTCAGCCCTGCTAGCCCGGAGCCCG 187  
QY 88 ATGGCTCAAGAGGCTGTGCGCGCCCTTCCCTCCAGAGGACGCCCTTAACCCCGCATCC 147  
DB 186 CCCCCAGAGGTGGCTGGCGCGCTCGAGGC-CCAGCGCGCGCGCGCGCGCGCGCGCC 128  
QY 148 CTGCGCCGCCACCCCTGTCTCCCGCTCTCGAGCCCTTTAATCTGACTTCGAGCCCTCG 207  
DB 127 GCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCTCCGCGCGCGCGCGCGCGCC 68  
QY 208 GACTAGCTCTGACCTCAGCACTTCTCTCCAGCAACCC 247  
DB 67 GCGCGCGCGCTCCGCC 28

## RESULT 15

US-08-584-226-3/c  
Sequence 3, Application US/08584226  
Patent No. 5798240

## GENERAL INFORMATION:

APPLICANT: Martinis, Susan A.  
APPLICANT: Sassanfar, Mandana  
APPLICANT: Kim, Sunghoon  
APPLICANT: Lee, Sang Ho  
APPLICANT: Schimmel, Paul R.  
TITLE OF INVENTION: RECOMBINANT MYCOBACTERIAL METHIONYL-TRNA  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington

STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173-4799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/584,226  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/305,766  
FILING DATE: 13-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: CPI94-052  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 599 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-584-226-3

Query Match 12.3%; Score 31.8; DB 1; Length 599;

Best Local Similarity 48.1%; Pred. No. 3;

Matches 90; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 65 CCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGCGCCCTTGCTCCAG 124  
DB 591 CCGCGGAATTCTATCTGACGGGCGGATGAGCAGCGCTGAAGGTGGCGAGGCGCGCG 532  
QY 125 AGGAGCGCCCTAACCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCGAGCCCT 184  
DB 531 CGCGCGCGCGGCTGCGCGACCGCACAGCTTGGCCCGCGCAATTCCGACGCTTCCACGCA 472  
QY 185 TTAATCTGACTTCGGAGCGCTCGGACTACGCTCTGGACTCAGACACTTCTCTCCAGCAAC 244  
DB 471 TGCAGGAGCGCTGNACATCTCTTCGACCGGTTTCATCCGACACCCCGGACCACT 412  
QY 245 ACCCGCG 251  
DB 411 ACAGCGC 405

Search completed: January 30, 2001, 21:28:34  
Job time: 19996 sec

APPLICANT: Weinzierl, Robert O.J.  
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/646,715  
FILING DATE: 09-MAY-1996  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/188,582  
FILING DATE: 28-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1120 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 80..913  
US-08-646-715-8

Query Match 13.0%; Score 33.6; DB 1; Length 1120;  
Best Local Similarity 51.0%; Pred. No. 1.2;  
Matches 104; Conservative 0; Mismatches 99; Indels 1; Gaps 1;  
QY 34 CAGCCAAAGGCTTCTTTCTGCTGCTCACCATGCTCTTGGCCAGATCTCTGATGCT 93  
DB 866 CCACCTCAACGCCAGATGATCTCTCTCCCTCCCTCTCTCCCTGCGCTGCGACATCTCCGCT-CCC 808  
QY 94 GAAGAGGCTGTCCCGCGCCCTGCTCCAGAGGAGCGCCCTTAACGCCATCCCTGGCG 153  
DB 807 GAACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 748  
QY 154 CCCACCCCTGTGCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 213  
DB 747 AGATCTCCCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 688  
QY 214 GCTCTGGACCTCAGCATTCTTCTC 237  
DB 687 CCGCCGGAGCTGCCACCGCTTTC 664

RESULT 13  
US-07-705-490-1/c  
Sequence 1, Application US/07705490  
Patent No. 6107025  
GENERAL INFORMATION:  
APPLICANT: Caskey, C. T.  
ADDRESSEE: Nelson, David L.  
APPLICANT: Pieretti, Maura

APPLICANT: Warren, Stephen T.  
ADDRESSEE: Oosiro, Ben A.  
TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Thomas D. Paul  
STREET: 1301 McKinney, Suite 5100  
CITY: Houston  
STATE: Texas  
COUNTRY: U.S.A.  
ZIP: 77010-3095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/705,490  
FILING DATE: 19910708  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul, Thomas D.  
REGISTRATION NUMBER: 32,714  
REFERENCE/DOCKET NUMBER: D-5350  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713/651-5325  
TELEFAX: 713/651-5246  
TELEX: 762829  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3765 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-07-705-490-1

Query Match 12.4%; Score 32; DB 3; Length 3765;  
Best Local Similarity 49.5%; Pred. No. 4.1;  
Matches 109; Conservative 0; Mismatches 110; Indels 1; Gaps 1;  
QY 28 CCGAACCCAGCCAAAGGCTTCTTTCTGCTGCTCACCATGCTCTTCTGCCAGATCTCTG 87  
DB 222 CCGCACCTTCACGACCCAGCT 163  
QY 88 ATGGCTGAAGAGGCTGTGCGCGCGCCCTGCTCTCCAGAGGAGCCCTTAACGCCATCTCC 147  
DB 162 CCGCCGAGAGGTGGCTGCGCGCGCTGAGGC-CCAGCCGCGCGCGCGCGCGCGCGCGCG 104  
QY 148 CTGGCGCCACCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 207  
DB 103 CCGCGCTCG 44  
QY 208 GACTAGCTGCTGAGCTCAGCATTCTTCTCCACCAACACC 247  
DB 43 CCGCGCGCGCTGCCGACGCCCTTGGCACGGCGCGCTCC 4

RESULT 14  
US-08-455-073A-1/c  
Sequence 1, Application US/08455073A  
Patent No. 5876949  
GENERAL INFORMATION:  
APPLICANT: Gideon Dreyfuss  
ADDRESSEE: Mikiko C. Sloni  
APPLICANT: Yan Zhang  
TITLE OF INVENTION: Fragile X Related Proteins, Compositions And Methods  
TITLE OF INVENTION: Of Making And Using The Same  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 23:09:03 ; Search time 21.17 Seconds  
(without alignments)  
275.836 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156  
Perfect score: 442  
Sequence: 1 VRROLVPEEPNPAKRLFL.....EPDYLALDLSTFLQHPAAF 86

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues  
Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR56:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	442	100.0	156	2	JC5537	differentiation-de
2	273	61.8	153	2	S33363	gly96 protein - mo
3	81.5	18.4	102	2	T33565	hypothetical prote
4	79.5	18.0	1596	2	A41216	guanine nucleotide
5	78	17.6	4957	2	T03455	ALR protein - huma
6	78	17.6	5262	2	T03454	ALR protein - huma
7	76	17.2	148	2	A81141	acyl CoA thioester
8	76	17.2	1677	2	T18344	p-glycoprotein E -
9	74	16.7	955	2	T00247	zinc finger protei
10	74	16.7	1561	2	T00248	zinc finger protei
11	73.5	16.6	272	2	A47021	pectic enzyme secr
12	73.5	16.6	925	2	S27920	nuclear antigen EB
13	73	16.5	1029	2	T30351	mucin-like protein
14	72.5	16.4	272	2	S28013	outC protein - Erw
15	71.5	16.2	626	1	NBH01A	platelet glycoprot
16	71	16.1	359	2	T33756	hypothetical prote
17	70.5	16.0	235	2	D83477	hypothetical prote
18	70.5	16.0	605	1	Q0B229	BRLF1 protein - hu
19	70	15.8	148	2	E81878	probable acyl-CoA
20	70	15.8	1233	2	T37045	nitrate reductase
21	69.5	15.7	1887	2	S61703	fatty-acid synthas
22	69	15.6	418	2	T19800	hypothetical prote
23	69	15.6	1402	2	I46707	translation initia
24	69	15.6	2187	2	T30826	nascent polypeptid
25	68.5	15.5	181	2	T08793	hypothetical prote
26	68.5	15.5	882	1	IJHUCE	cadherin 1 precurs
27	68.5	15.5	1072	2	T50949	verprolin related
28	68.5	15.5	1337	1	I38670	protein-tyrosine-p
29	68.5	15.5	1844	2	S01956	hypothetical prote

30 68.5 15.5 2649 2 T51023 hypothetical prote  
31 68 15.4 241 2 T22216 hypothetical prote  
32 68 15.4 267 2 A49068 craniostomatosis-as  
33 68 15.4 1396 2 A44453 translation initia  
34 67 15.2 283 2 S13383 hydroxyproline-ric  
35 67 15.2 586 2 S19381 hypothetical prote  
36 67 15.2 3566 1 A40701 tenascin-X precurs  
37 66.5 15.0 169 2 A72466 hypothetical prote  
38 66.5 15.0 3149 1 Q0B28 BPLF1 protein - hu  
39 66 14.9 101 2 T46506 hypothetical prote  
40 66 14.9 187 2 C75558 acetyl-CoA carboxy  
41 66 14.9 621 2 JG5164 acetolactate synth  
42 66 14.9 5149 2 F83345 probable modR bos  
43 65.5 14.8 325 2 D70666 probable modR bos  
44 65.5 14.8 329 2 S28442 exdB protein - Pse  
45 65.5 14.8 424 2 T43468 hypothetical prote

ALIGNMENTS

RESULT 1  
JC5537  
differentiation-dependent protein DIF-2 - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Sep-1997 #sequence\_revision 05-Sep-1997 #text\_change 07-May-1999  
C:Accession: JC5537  
R:Pietzsch, A.; Buechler, C.; Aslanidis, C.; Schmitz, G.  
Biochem. Biophys. Res. Commun. 235, 4-9, 1997  
A:Title: Identification and characterization of a novel monocyte/macrophage differ.  
A:Reference number: JC5537; MUID:97339426  
A:Accession: JC5537  
A:Molecule type: mRNA  
A:Residues: 1-156 <PIE>  
A:Experimental source: monocyte  
A:Note: the authors translated the codon CCG for residue 106 as Arg

Query Match 100.0%; Score 442; DB 2; Length 156;  
Best Local Similarity 100.0%; Pred. No. 6.2e-36;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VRROLVPEEPNPAKRLFLLLTIVFCQLMAEEGVPAFLPPEDAPNAASLAPTPVSPVLE 60  
|||||  
Db 71 VRROLVPEEPNPAKRLFLLLTIVFCQLMAEEGVPAFLPPEDAPNAASLAPTPVSPVLE 130  
Qy 61 PFNLTSEPSDYALDLSTFLQHPAAF 86  
|||||  
Db 131 PFNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 2  
S33363  
gly96 protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Feb-1997  
C:Accession: S33363  
R:Charles, C.H.; Yoon, J.K.; Simske, J.S.; Lau, L.F.  
Oncogene 8, 797-801, 1993  
A:Title: Genomic structure, cDNA sequence, and expression of gly96, a growth factor  
A:Reference number: S33363; MUID:93173526  
A:Accession: S33363  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-153 <CHA>  
A:Cross-references: EMBL:X67644  
C:Genetics:  
A:Introns: 70/3  
C:Keywords: transmembrane protein  
Query Match 61.8%; Score 273; DB 2; Length 153;  
Best Local Similarity 68.2%; Pred. No. 1.3e-19;

[illegible]

A:Reference number: A94173; MUID:87289654  
 A:Accession: A94173  
 A:Molecule type: protein  
 A:Residues: 17-315 <TIT>  
 R:Hess, D.; Schaller, J.; Rickli, E.E.; Clemetson, K.J.  
 Eur. J. Biochem. 199, 389-393, 1991  
 A:Title: Identification of the disulphide bonds in human platelet glycosylcalicin.  
 A:Reference number: S16945; MUID:91301149  
 A:Accession: S16945  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 224-227;262-270;277-282 <HES>  
 R:Lopez, J.A.; Ludwig, E.H.; McCarthy, B.J.  
 J. Biol. Chem. 267, 10055-10061, 1992  
 A:Title: Polymorphism of human glycoprotein Ib alpha results from a variable number of t  
 ations.  
 A:Reference number: I55355; MUID:92250564  
 A:Accession: I55355  
 A:Status: preliminary; translated from GB/EMBL/NDJ  
 A:Molecule type: DNA  
 A:Residues: 412-427 <RES>  
 A:Cross-references: GB:S34436; NID:9249176; PIDN:AB22152.1; PID:9249177  
 A:Note: variant D  
 C:Comment: Glycoprotein Ib (GP1b), a surface membrane protein of platelets, participates  
 C:Comment: Platelet activation apparently involves disruption of the macromolecular comp  
 C:Comment: Binding sites for von Willebrand factor and thrombin (the latter site with un  
 C:Comment: Glycosylcalicin, which is approximately coextensive with the extracellular part  
 C:Genetics:  
 A:Gene: GDB:GP1BA; GPLB  
 A:Cross-references: GDB:I18806; OMIM:231200  
 A:Map position: 17pter-17p12  
 C:Complex: heterodimer with platelet glycoprotein Ib beta chain (NBHUIB)  
 C:Superfamily: platelet glycoprotein Ib alpha chain; leucine-rich alpha-2-glycoprotein x  
 C:Keywords: blood coagulation; duplication; glycoprotein; platelet membrane; tandem repe  
 F:1-16/Domain: signal sequence #status predicted <SIG>  
 F:17-626/Product: platelet glycoprotein Ib alpha chain #status predicted <MPT>  
 F:48-71/Domain: leucine-rich alpha-2-glycoprotein repeat homolog <LRR1>  
 F:72-93/Domain: leucine-rich alpha-2-glycoprotein repeat homolog <LRR2>  
 F:94-116/Domain: leucine-rich alpha-2-glycoprotein repeat homolog <LRR3>  
 F:117-140/Domain: leucine-rich alpha-2-glycoprotein repeat homolog <LRR4>  
 F:141-164/Domain: leucine-rich alpha-2-glycoprotein repeat homolog <LRR5>  
 F:165-188/Domain: leucine-rich alpha-2-glycoprotein repeat homolog <LRR6>  
 F:189-212/Domain: leucine-rich alpha-2-glycoprotein repeat homolog <LRR7>  
 F:379-430/Region: proline/threonine-rich 9-residue repeats  
 F:502-540/Domain: transmembrane #status predicted <TRM>  
 F:541-626/Domain: intracellular #status predicted <INT>  
 F:37.175/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:308/Binding site: carbohydrate (Thr) (covalent) #status experimental

Query Match 16.2%; Score 71.5; DB 1; Length 626;  
 Best Local Similarity 45.0%; Pred. No. 17;  
 Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;  
 Oy 36 PAPLP-----PEDAPNAASLAPTPVSPVLEPFLNLTSEPS 69  
 |||||  
 Db 382 PTPPTTSEPVPPEPAPNNTLTLEPTPTPEP---TSEPA 418

Search completed: January 30, 2001, 23:09:04  
 Job time: 6029 sec



```

Oy 54 --PVSPVLEPF--NLTSEP-----SDYALDLSTFLQO----HPAAF 86
      |:::| |:::| |:::| |:::| |:::| |:::| |:::|
Db 1478 TSPLTPAMSPNPINPSPVSTSSSYAHQLRNRQOQOQTTPAIR 1523

```

RESULT 5  
 US-07-821-717B-6  
 ; Sequence 6, Application US/0782171B  
 ; Patent No. 5298239  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Miller, Jonathan L.  
 ; APPLICANT: Cunningham, David  
 ; APPLICANT: Lyle, Vicki A.  
 ; APPLICANT: Finch, Clara N.  
 ; TITLE OF INVENTION: MUTATIONS RENDERING PLATELET  
 ; TITLE OF INVENTION: GLYCOPROTEIN Ib ALPHA LESS REACTIVE  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
 ; STREET: Clinton Square, P.O. Box 1051  
 ; CITY: Rochester  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 14603  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/821,717B  
 ; FILING DATE: 15-JAN-1992  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Timain, Susan J  
 ; REGISTRATION NUMBER: 34,103  
 ; REFERENCE/DOCKET NUMBER: 20884/21  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (716) 263-1636  
 ; TELEFAX: (716) 263-1600  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 610 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; PUBLICATION INFORMATION:  
 ; AUTHORS: Lopez, Jose A.  
 ; AUTHORS: Chung, Dominic W.  
 ; AUTHORS: Fujikawa, Kazuo  
 ; AUTHORS: Hagen, Frederick S.  
 ; AUTHORS: Papayannopoulou, Thalia  
 ; AUTHORS: Roth, Gerald J.  
 ; TITLE: Cloning of the alpha chain of human  
 ; TITLE: platelet glycoprotein Ib: A transmembrane prote  
 ; TITLE: to leucine-rich alpha-2-glycoprotein  
 ; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.  
 ; VOLUME: 84  
 ; PAGES: 5615-5619  
 ; DATE: AUG-1987  
 ; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 610  
 ; PUBLICATION INFORMATION:  
 ; AUTHORS: Zimmerman, Theodore S.  
 ; AUTHORS: Ruggeri, Zaverio M.  
 ; AUTHORS: Houghten, Richard A.  
 ; AUTHORS: Vincete, Vincete  
 ; AUTHORS: Mohri, Hiroshi  
 ; TITLE: Peptidic fragments and synthetic  
 ; TITLE: peptides that block the binding of von Willebra  
 ; TITLE: platelet membrane glycoprotein Ib  
 ; DOCUMENT NUMBER: EP 0 317 278 A2

```

: FILING DATE: 16-NOV-1988
: PUBLICATION DATE: 24-MAY-1989
: RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 293
US-07-821-717B-6

Query Match      16.2%; Score 71.5; DB 1; Length 610;
Best Local Similarity 45.08; Pred. NO. 2.7;
Matches 18; Conservative 2; Mismatches 11; Indels

QY 36 PAPLP-----PEDAPNAASLAPTPVSPVLEPFLNLTSEPS 69
      ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 366 PTPSPTTSEPVPEPAPNNLTLEPTPSTTPEP-----TSEPÀ 402

RESULT 6
US-08-119-262B-6
: Sequence 6, Application US/08119262B
: Patent No. 5492809
: GENERAL INFORMATION:
: APPLICANT: Miller, Jonathan L.
: APPLICANT: Cunningham, David
: APPLICANT: Lyle, Vicki A.
: APPLICANT: Finch, Clara A.
: TITLE OF INVENTION: MUTATIONS RENDERING PLATELET
: TITLE OF INVENTION: GLYCOPROTEIN IB ALPHA LESS REACTIVE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nixon, Hargrave, Devans & Doyle
: STREET: Clinton Square, P.O. Box 1051
: CITY: Rochester
: STATE: New York
: COUNTRY: USA
: ZIP: 14603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/119,262B
: FILING DATE: 09-SEP-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/821,717
: FILING DATE: 15-JAN-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Timain, Susan J.
: REGISTRATION NUMBER: 34,103
: REFERENCE/DOCKET NUMBER: 20884/22
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (716) 263-1636
: TELEFAX: (716) 263-1600
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 610 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PUBLICATION INFORMATION:
: AUTHORS: Lopez, Jose A.
: AUTHORS: Chung, Dominic W.
: AUTHORS: Fujikawa, Kazuo
: AUTHORS: Hagen, Frederick S.
: AUTHORS: Papayannopoulou, Thalia
: AUTHORS: Roth, Gerald J.
: TITLE: Cloning of the alpha chain of human platelet
: TITLE: glycoprotein Ib: A transmembrane protein
: TITLE: leucine-rich alpha-2-glycoprotein
: JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
: VOLUME: 84
: PAGES: 5615-5619

```

with hon





ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 22640720  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 286 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
US-08-382-184-3

Query Match 14.8%; Score 65.5; DB 1; Length 286;  
Best Local Similarity 46.9%; Pred. NO. 5;  
Matches 15; Conservative 3; Mismatches 11; Indels

**Qy** 29 LMAEEGVAPLPPEDAP---NAASLAPTVPSP 57  
| | ||| { | | | : |||| : |  
**Dd** 247 LVAPPAPAPAAEPAPAPAPAGEVAPTPTTP 278

RESULT 15

US-08-641-356-3  
Sequence 3, Application US/08641356  
Patent No. 5866130  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: MICROBACTERIAL PROTEINS.  
TITLE OF INVENTION: MICROORGANISMS PRODUCING THEM AND  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/641.356  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/382.184  
FILING DATE: 01-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 22640720  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 286 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
US-08-641-356-3

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 22:05:37 ; Search time 54.97 Seconds  
(without alignments)  
53.496 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156

Perfect score: 442

Sequence: 1 VRQLPVEENPAKRLFL.....EPDYLALDLSTFLQHPAAF 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_36:\*\*

1: /cgn2\_2/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /cgn2\_2/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /cgn2\_2/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /cgn2\_2/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /cgn2\_2/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /cgn2\_2/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /cgn2\_2/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /cgn2\_2/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /cgn2\_2/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /cgn2\_2/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /cgn2\_2/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /cgn2\_2/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /cgn2\_2/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /cgn2\_2/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /cgn2\_2/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /cgn2\_2/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /cgn2\_2/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /cgn2\_2/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /cgn2\_2/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /cgn2\_2/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /cgn2\_2/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	442	100.0	156	18 W36006	Human Fchd605 gene
2	442	100.0	156	21 Y45017	Protein encoded by
3	79.5	18.0	1596	21 Y68821	Amino acid sequenc
4	71.5	16.2	610	15 R51116	Platelet glycoprot
5	71.5	16.2	610	15 R56664	Mutant platelet g
6	71.5	16.2	610	17 R89436	Mutated platelet g
7	71.5	16.2	610	18 W18201	Platelet glycoprot
8	69.5	15.7	739	18 W25790	Gene 036 product d
9	68.5	15.5	204	18 W14578	Streptococcus pneu
10	68.5	15.5	370	21 Y99428	Human PRO1431 (UNQ
11	68.5	15.5	878	15 R55060	Sequence of human
12	68.5	15.5	878	16 R85487	Human E-cadherin p

13	68.5	15.5	882	20 Y09375	Wild-type E-cadher
14	68.5	15.5	1337	16 R85203	huDEP-1. Homo sap
15	67	15.2	541	19 W37148	Mammalian Ena (Men
16	67	15.2	783	19 W37151	Mouse neural Mena+
17	67	15.2	787	19 W37152	Mouse neural Mena+
18	67	15.2	802	19 W37153	Mouse neural Mena+
19	66.5	15.0	131	18 W27646	Secreted protein A
20	66.5	15.0	131	18 W44082	Human secreted pro
21	65.5	14.8	128	18 W14577	Streptococcus pneu
22	65.5	14.8	286	17 W03566	Mycobacterium tube
23	65.5	14.8	325	17 W03565	Mycobacterium tube
24	65.5	14.8	332	18 W32418	Mycobacterium tube
25	65.5	14.8	332	18 W32350	Mycobacterium tube
26	65.5	14.8	332	19 W81683	M. tuberculosis im
27	65.5	14.8	332	19 W64322	Mycobacterium tube
28	65.5	14.8	332	20 Y39083	M. tuberculosis an
29	65.5	14.8	332	20 Y38945	M. tuberculosis re
30	65.5	14.8	456	20 Y17067	Human 3-OST-4 prot
31	65.5	14.8	652	20 Y39225	M. tuberculosis fu
32	65.5	14.8	652	20 Y39082	M. tuberculosis fus
33	65.5	14.8	802	19 W81746	M. tuberculosis fu
34	65.5	14.8	802	19 W64379	Mycobacterium anti
35	65.5	14.8	802	20 Y32063	Mycobacterium tube
36	65.5	14.8	802	20 Y39224	M. tuberculosis fu
37	65.5	14.8	802	20 Y39176	M. tuberculosis fu
38	65.5	14.8	802	20 Y39081	M. tuberculosis fus
39	65.5	14.8	802	20 Y39033	M. tuberculosis fu
40	65.5	14.8	1093	16 R66460	AF-17 protein. Ho
41	65	14.7	328	19 W36984	E. coli zipA prote
42	65	14.7	427	21 Y68748	Amino acid sequenc
43	64.5	14.6	147	15 R60799	Rape abscission/de
44	64.5	14.6	194	18 W14584	Streptococcus pneu
45	64.5	14.6	412	20 Y49151	Amino acid sequenc

#### ALIGNMENTS

#### RESULT 1

ID W36006 standard; Protein: 156 AA.

AC W36006;

DT 03-MAR-1998 (first entry)

DE Human Fchd605 gene product.

XX Fchd605 gene: differential expression; monocyte; human;  
KW foam cell; cardiovascular disease; atherosclerosis; ischaemia;  
KW reperfusion; hypertension; restenosis; arterial inflammation;  
KW therapy; diagnosis; drug screening; marker.

OS Homo sapiens.

PN W09730065-A1.

PD 21-AUG-1997.

XX 14-FEB-1997: 97WO-US02291.

PR 13-FEB-1997: 97US-0799910.

PR 16-FEB-1996: 96US-0011787.

PA (MILL-) MILLENNIUM PHARM INC.

XX Falb DA;

WPI: 1997-424966/39.

DR N-PSDB; T94471.

XX New genes differentially expressed in cardiovascular disease - used  
PT for diagnosis, drug screening and treatment of cardiovascular

A substitution in platelet glycoprotein Ib alpha (Leucine 57 to CC phenylalanine) underlies a form of Bernard-Soulier disease. The CC mutated glycoprotein can be used in compositions to inhibit CC platelet aggregation/adhesion. The glycoprotein may be labelled and CC used as an imaging agent and may also be bound to a thrombolytic CC agent, preferably tissue plasminogen activator (tPA), CC (pro) urokinase, streptokinase, anisoylated plasminogen-streptokinase CC activator complex, tPA analogues or a protease, allowing localisation CC of the thrombolytic agent to a thrombus.

Qy 36 PAPLP-----PEDAPNAASLATPVSPVLEFFNLTSPPS 69  
| | | | | : | | | |  
Db 366 ptpsppttsevppeapamnttleptcpsptltppep---tsapa 40

07545M

XX  
195420;



QY 21 LTVFCQILMA-----EGVPAPLPPEDAPNAASLAPTPVSPVLEPNLTSEP-SDVAL 73.  
 Db 793 lttvscgkmaaptrtcttgtdpppgspnitsvshsvkxkfsgfeashgpkikayav 852  
 QY 74 DLSTFLQHPAA 85  
 Db 853 ltttgeaghpsa 864

RESULT 15  
 W37148  
 ID W37148 standard; Protein: 541 AA.  
 AC W37148;  
 XX  
 DT 06-JUL-1998 (first entry)  
 XX  
 DE Mammalian Ena (Mena).  
 XX  
 KW Mena protein; mammalian Ena; Enabled; Evi protein; cytoskeleton;  
 KW cell morphology; cell adhesion; cell differentiation; cell growth;  
 KW cell motility; mouse.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 1..113  
 FT Peptide 156..160  
 FT Peptide 166..170  
 FT Peptide 171..175  
 FT Peptide 176..180  
 FT Peptide 199..203  
 FT Peptide 207..211  
 FT Peptide 236  
 FT Modified-site /note= "O-phosphorylated"  
 FT Binding-site 305..310  
 FT Binding-site /note= "mediates binding to profilin"  
 FT Binding-site 316..321  
 FT Binding-site /note= "mediates binding to profilin"  
 XX  
 PN W09801755-A1.  
 XX  
 PD 15-JAN-1998.  
 XX  
 PF 03-JUL-1997; 97WO-US11669.  
 XX  
 PR 05-JUL-1996; 96US-0675815.  
 XX  
 PA (GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.  
 PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
 XX  
 PI Gertler FB, Niebuhr K, Soriano P, Wehland J;  
 XX  
 DR WPI: 1998-101197/09.  
 DR N-PSDB: V02996.  
 XX  
 PT Detection of modulators of Mena and Ena-VASP-like genes and proteins  
 PT - used in control of cytoskeletal dynamic events in normal and  
 PT abnormal cell morphology, adhesion, motility, growth and  
 PT differentiation  
 XX  
 PS Example 1: Page 54-56; 77pp; English.  
 XX  
 CC This polypeptide comprises murine Mena (mammalian Ena) that shows  
 CC significant amino acid similarity to Drosophila Ena and which  
 CC exhibits a broad pattern of tissue distribution in neurons,

CC fibroblasts, kidney epithelium, muscle, neural crest and  
 CC hematopoietic cells. Its amino acid sequence was deduced from  
 CC a cDNA clone (see V02996) obtained from a mouse embryonic stem cell  
 CC cDNA library. 2 Novel mammalian genes, Mena and Ena-VASP-like  
 CC (Evl, see V02997), encoding novel proteins Mena and Evi (see  
 CC W37149) are disclosed. Mena and Evi proteins have a discrete, EVH1  
 CC functional domain responsible for Mena binding to Listeria, and to  
 CC the cytoskeletal proteins zyxin and vinculin. The EVH1 domain of  
 CC Mena is also responsible and sufficient for targeting localisation  
 CC of Mena and Mena-based fusion proteins to focal adhesions, and to  
 CC the surface of Listeria cells at the polar site of induction of  
 CC actin comet tail formation. Based on the disclosed Mena and Evi  
 CC genes and proteins, a variety of methods and compositions are  
 CC provided for screening, isolating and characterising endogenous and  
 CC exogenous factors, drugs and therapeutic agents useful to evaluate  
 CC and/or control cytoskeletal dynamic events involved in normal and  
 CC abnormal cell morphology, adhesion, motility, growth and/or  
 CC differentiation. A method of detecting a modulator of Mena  
 CC activity/expression is claimed.

XX  
 SQ Sequence 541 AA;

Query Match 15.2%; Score 67; DB 19; Length 541;  
 Best Local Similarity 29.9%; Pred. No. 25;

Matches 20: Conservative 4; Mismatches 23; Indels 20; Gaps 2;

QY 6 PVEEPNPAKRLFLLLTLIVFCQILMAEGVPAPLPPEDAPN-----AASLAPTPVSPVLEP 61  
 Db 268 paesptpggvlv-----gppapppppplpsgpayasalpppppppppp 311

QY 62 FNLTSEP 68  
 Db 312 lpstgpp 318

Search completed: January 30, 2001, 22:05:37  
 Job time: 11886 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 22:04:55 ; Search time 64.84 Seconds  
(without alignments)  
281.993 Million cell updates/sec

Title: US-08-799-910-10  
Perfect score: 823  
Sequence: 1 MCHSRCHPTMTILOAPTPA.....EPSDYALDLSTFLOQHPAAF 156

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_15.\*

- 1: sp-archaea.\*
- 2: sp-bacteria.\*
- 3: sp-fungi.\*
- 4: sp-human.\*
- 5: sp-invertebrate.\*
- 6: sp-mammal.\*
- 7: sp-mhc.\*
- 8: sp-organelle.\*
- 9: sp-phage.\*
- 10: sp-plant.\*
- 11: sp-rodent.\*
- 12: sp-virus.\*
- 13: sp-vertebrate.\*
- 14: sp-unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	771.5	93.7	193	4	09Y4F5
2	122.5	14.9	423	4	09Y4F5
3	114.5	13.9	1668	4	09Y4F5
4	114.5	13.9	2971	4	09Y4F5
5	114	13.9	3247	12	065553
6	112.5	13.7	649	5	09V884
7	110.5	13.4	1011	3	09P844
8	109.5	13.3	552	10	09G343
9	109.5	13.3	1315	10	09SPM0
10	109	13.2	285	3	P78977
11	109	13.2	1709	4	015047
12	109	13.2	3325	12	09L8T9
13	108.5	13.2	233	2	09L568
14	107.5	13.1	203	2	09RCX9
15	107.5	13.1	581	5	020517
16	107.5	13.1	1013	5	09VY77
17	107	13.0	426	2	09LAY5
18	107	13.0	496	12	098457
19	106.5	12.9	901	2	044562

20	106.5	12.9	1498	4	09Y4F5
21	106	12.9	1029	12	09YMX0
22	106	12.9	2066	12	P89920
23	105.5	12.8	249	2	09L570
24	105	12.8	296	5	09V729
25	105	12.8	351	10	039492
26	105	12.8	549	12	089370
27	104.5	12.7	243	2	09L567
28	104.5	12.7	244	2	09L565
29	104.5	12.7	535	2	P74745
30	104.5	12.7	616	12	096716
31	104	12.6	395	2	09LAY2
32	104	12.6	408	2	09LAY0
33	103	12.5	241	5	045522
34	102.5	12.5	288	12	084565
35	102.5	12.5	767	5	062537
36	102	12.4	336	2	09L569
37	102	12.4	331	5	09V728
38	102	12.4	708	10	09SX31
39	101.5	12.3	344	11	089037
40	101	12.3	339	12	084465
41	101	12.3	395	2	09LAZ1
42	101	12.3	437	2	09LAY4
43	101	12.3	739	2	09ROT4
44	101	12.3	852	4	09NP71
45	100	12.2	802	11	P70433

## ALIGNMENTS

RESULT 1

075353 ID O75353 PRELIMINARY; PRT; 193 AA.  
AC O75353;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE ANTI-DEATH PROTEIN.  
GN IEX-1L.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98369175; PubMed=9703517;  
RA Wu M.X., Ao Z., Prasad K.V.S., Wu R., Schlossman S.F.;  
RT "IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival.";  
RL Science 281:998-1001(1998).  
DR EMBL; AF039067; AAC32558.1;  
DR EMBL; AF071596; AAC72344.1;  
SQ SEQUENCE 193 AA; 7927D9D3FFBC7C57 CRC64;

Query Match 93.7%; Score 771.5; DB 4; Length 193;  
Best Local Similarity 78.8%; Pred. No. 8.1e-64;  
Matches 152; Conservative 1; Mismatches 3; Indels 37; Gaps 1;

QY	1	MCHSRCHPTMTILOAPTPAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASGRHK 60		
DB	1	MCHSRCHPTMTILOAPTPAPSTIPGPRGSGPEIFTDPLPEPAAAPAGRPSASGRHK 60		
QY	61	RSRRVLYPR-----VVRQLPVEENPA 83		
DB	61	RSRRVLYPR-----VVRQLPVEENPA 120		
QY	84	KRLLELLTIVFCOILMAEGVPAPLPEDAPNAASLAPTPVSPVLEPFNLTSPSYAL 143		
DB	121	KRLLELLTIVFCOILMAEGVPAPLPEDAPNAASLAPTPVSPVLEPFNLTSPSYAL 180		
QY	144	DLSTFLQHPAAF 156		





RX MEDLINE-97127825; PubMed-8972576;  
 RA Ramon A., Gil R., Burgal M., Sentandreu R., Valentin E.;  
 RT "A novel cell wall protein specific to the mycelial form of *Yarrowia*  
 RT *lipolytica*."  
 RL Yeast 12:1535-1548(1996).  
 DR EMBL: Z81006; CAB02634.1;  
 KW Signal.  
 FT SIGNAL 1 16 POTENTIAL.  
 FT CHAIN 17 285 POTENTIAL.  
 SQ SEQUENCE 285 AA; 30036 MW; OBA7FF1C515CD57D CRC64;

Query Match 13.2%; Score 109; DB 3; Length 285;  
 Best Local Similarity 28.4%; Pred. No. 0.012;  
 Matches 40; Conservative 11; Mismatches 62; Indels 28; Gaps 6;

QY 1 MCHSRCHPTMTILQATTPAP-STIPGRGSGPEIFTDPLPEAAAPAGPSASRGHR 59  
 DB 122 ICTKTVLSTVLTPTTPVPTTTPAVEPKPTPEVVKPEPTP-EVGVKPEPTRGPP 180  
 QY 60 KRSRRVLYPRVVRNOLP-VEEPNPAKRLLFLLLTIVFCQILMAEGVPAPLPDPEDAPNAA 118  
 DB 181 AKPEPEVEVKPEPTPEVVRP-----EPTTAPLPPTP-PSLR 218  
 QY 119 SLAPTPVSPVLEPFLNLTSEPS 139  
 DB 219 SLRS---SPSLPPLPLPSPS 236

RESULT 11  
 Q15047 PRELIMINARY; PRT; 1709 AA.  
 ID O15047;  
 AC O15047;  
 DT 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)  
 DE K1A0339 PROTEIN (FRAGMENT).  
 GN K1A0339.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro."  
 RL DNA Res. 4:141-150(1997).  
 DR EMBL: AB002337; BAA20797.2;  
 DR INTERPRO: IPR000504;  
 DR INTERPRO: IPR001214;  
 DR INTERPRO: IPR002965;  
 DR PFAM: PF00076; trm. 1;  
 DR PFAM: PF00856; SET. 1;  
 DR PRINTS: PR01217; PRICHEXTNSN.  
 FT NON\_TER 1  
 SQ SEQUENCE 1709 AA; 186231 MW; DCF2E0FF716B672C CRC64;

Query Match 13.2%; Score 109; DB 4; Length 1709;  
 Best Local Similarity 32.6%; Pred. No. 0.069;  
 Matches 44; Conservative 10; Mismatches 53; Indels 28; Gaps 7;

QY 6 SCHPTMTILQATTPAPSTIPGRGSGPEIFTDPLPEAAAPAGPSASRGHRKRRV 65  
 DB 1078 SASPPPEVPVPTAPVEVPVVRVAGSPV---TLPQEASPA-RVAGTTEESPPSAPL 1133  
 QY 66 LYPRVVRQLPVEEPNPAKR-----LLFLL-----LTIVFCQILMAEGVPAPLP 111  
 DB 1134 RPP-----EPPAGPPAPAPRDERPSSPIPLPPKRRKRTVTSFSAI-----EVVPAPEPP 1184

QY 112 EDAPNAASLAPTPVS 126  
 DB 1185 PATPPQAKF-PCPAS 1198

RESULT 12  
 Q91BT9 PRELIMINARY; PRT; 3325 AA.  
 ID Q91BT9;  
 AC Q91BT9;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)  
 DE MAJOR TEGUMENT PROTEIN.  
 GN UL36.  
 OS Turkey herpesvirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.  
 OX NCBI\_TaxID=10390;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GA;  
 RX MEDLINE=92237304; PubMed=1315048;  
 RA Lee D., Lee L., Liu J.L., Kung H.J., Tillotson J.K.;  
 RT "Varek disease virus encodes a basic-leucine zipper gene resembling  
 RT the fos/jun oncogenes that is highly expressed in lymphoblastoid  
 RT tumors."  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4042-4046(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GA;  
 RA Lee L.F., Wu P., Sui D., Ren D., Kung H.J.;  
 RT "The Complete UL Sequence of Serotype I Marek's Disease Virus."  
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2000).  
 DR EMBL: AF147806; AAF66771.1;  
 SQ SEQUENCE 3325 AA; 365698 MW; 5D4A84EF719B99FD CRC64;

Query Match 13.2%; Score 109; DB 12; Length 3325;  
 Best Local Similarity 25.2%; Pred. No. 0.13;  
 Matches 31; Conservative 14; Mismatches 52; Indels 26; Gaps 3;  
 QY 16 APTPAPSTIPGRGSGPEIFTDPLPEAAAPAGPSASRGHRKRRVLYPRVVRQL 75  
 DB 2802 SPAPKSPAPKPPPPDPD---FKPSPAPKSPASKPSASKPSASKPPAPDSKPS 2858  
 QY 76 PVEEPNPAKRLLFLLLTIVFCQILMAEGVPAPLPDPEDAPNAASLAPTP-----VSPV 128  
 DB 2859 PAPKPPPP-----TPDSKPSAPKPSASKPSASKPLPVVPFPFNSDSKTSVP 2902  
 QY 129 LEP 131  
 DB 2903 PNP 2905

RESULT 13  
 Q9L568 PRELIMINARY; PRT; 233 AA.  
 ID Q9L568;  
 AC Q9L568;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)  
 DE PSPA (FRAGMENT).  
 GN PSPA.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-39;  
 RA Beall B.W.;  
 RT "PSPA sequence types from multiresistant pneumococci."  
 RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 21:28:35 ; Search time 21.17 seconds  
(without alignments)  
500.354 Million cell updates/sec

Title: US-08-799-910-10  
Perfect score: 823  
Sequence: 1 MCHSRCHPTMTILQAPTPA.....EPSDYALDLSTFLQHPAAF 156

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: PIR\_66:\*

1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	819	99.5	156	2	JC5537	differentiation-de
2	525	63.8	153	2	S33363	gly96 protein - mo
3	112.5	13.7	283	2	S13383	hydroxyproline-ric
4	109.5	13.3	552	2	T08148	proline-rich myros
5	107.5	13.1	581	2	T22341	hypothetical prote
6	107	13.0	496	2	T17908	proline/lysine-ric
7	106.5	12.9	901	2	A49227	siyalidase - Actino
8	106.5	12.9	1323	2	T00037	hypothetical prote
9	106	12.9	1029	2	T30351	mucin-like protein
10	105	12.8	351	2	S50754	hypothetical prote
11	105	12.8	549	2	T17525	proline-rich prote
12	105	12.8	3149	1	Q8E8	BPLF1 protein - hu
13	104.5	12.7	535	1	S76953	protein kinase (BC
14	104	12.6	815	2	B56708	extracellular sign
15	103	12.5	241	2	T22216	hypothetical prote
16	102.5	12.5	288	2	T17737	proline-rich prote
17	102.5	12.5	3866	2	B48205	All-1 protein -GPE
18	102.5	12.5	3869	2	A48205	All-1 protein +GTE
19	101	12.3	339	2	T17636	proline-rich prote
20	100	12.2	620	2	S06733	hydroxyproline-ric
21	99.5	12.1	279	2	T10361	hypothetical prote
22	98.5	12.0	3164	1	WMBEH6	UL36 protein - hum
23	98	11.9	473	2	S50755	hypothetical prote
24	97.5	11.8	225	2	T17815	proline-rich prote
25	97.5	11.8	2157	2	S71461	proline-rich prote
26	97.5	11.8	3338	2	T42761	Bassoon protein -
27	97	11.8	461	2	T10741	extensin-like prot
28	96.5	11.7	168	2	S52994	arabinogalactan-li
29	96	11.7	185	2	T00519	proline-rich prote

30 95.5 11.6 524 2 A75588 probable protein k  
31 95.5 11.6 537 2 A46611 myosin-binding pro  
32 95.5 11.6 753 2 JQ0532 OP protein - Kanne  
33 95.5 11.6 1596 2 A41216 guanine nucleotide  
34 95 11.5 202 2 JQ0964 hydroxyproline-ric  
35 95 11.5 377 2 A48018 mucin 7 precursor,  
36 95 11.5 416 1 SKXLAG dermal gland prote  
37 95 11.5 801 2 T29018 hypothetical prote  
38 95 11.5 3511 2 A59295 unconventional myo  
39 94.5 11.5 510 1 S43692 transcription fact  
40 94.5 11.5 758 2 A29253 finger protein hun  
41 94.5 11.5 2342 2 T13412 hypothetical prote  
42 94 11.4 328 2 JQ0985 hydroxyproline-ric  
43 94 11.4 925 2 S27920 nuclear antigen EB  
44 94 11.4 1290 2 T00018 period protein hom  
45 93.5 11.4 316 2 T31880 hypothetical prote

## ALIGNMENTS

### RESULT 1

JC5537

differentiation-dependent protein DIF-2 - human

C:Species: Homo sapiens (man)

C>Date: 02-Sep-1997 #sequence\_revision 05-Sep-1997 #text\_change 07-May-1999

C:Accession: JC5537

R:Pietzsch, A.; Buechler, C.; Aslanidis, C.; Schmitz, G.

Biochem. Biophys. Res. Commun. 235, 4-9; 1997

A>Title: Identification and characterization of a novel monocyte/macrophage differ

A:Reference number: JC5537; MUID:97339426

A:Accession: JC5537

A:Molecule type: mRNA

A:Residues: 1-156 <PIF>

A:Experimental source: monocyte

A>Note: the authors translated the codon CCG for residue 106 as Arg

Query Match 99.5%; Score 819; DB 2; Length 156;  
Best Local Similarity 99.4%; Pred. No. 1.2e-58;  
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MCHSRCHPTMTILQAPTPAPSTIPGPRRGSGPEITFDLPPEPAAAPGRPSASGRHK 60  
|||||  
DB 1 MCHSRCHPTMTILQAPTPAPSTIPGPRRGSGPEITFDLPPEPAAAPGRPSASGRHK 60  
QY 61 RRRVLYPRVRRQLPVEEPNPAKRLFLLLTVFCQILMAEGVPAPLPDPAPNAASL 120  
|||||  
DB 61 RRRVLYPRVRRQLPVEEPNPAKRLFLLLTVFCQILMAEGVPAPLPDPAPNAASL 120  
QY 121 APTVPSPVLEPFLNLTSEPSDYALDLSTFLQHPAAF 156  
|||||  
DB 121 APTVPSPVLEPFLNLTSEPSDYALDLSTFLQHPAAF 156

### RESULT 2

S33363

gly96 protein - mouse

C:Species: Mus musculus (house mouse)

C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Feb-1997

C:Accession: S33363

R.Charles, C.H.; Ioon, J.K.; Sinske, J.S.; Lau, L.F.

Oncogene 8, 797-801, 1993

A:Title: Genomic structure, cDNA sequence, and expression of gly96, a growth factor:

A:Reference number: S33363; MUID:93173526

A:Accession: S33363

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-153 <CHA>

A:Cross-references: EMBL:X67644

C:Genetics:

A:Introns: 70/3

C:Keywords: transmembrane protein





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 22:05:39 ; Search time 43.61 seconds  
(without alignments)  
115.521 Million cell updates/sec

Title: US-08-799-910-10

Perfect score: 823

Sequence: 1 MCHSRCHPTMTILQAPTPA.....EPSDYALDLSTFLQHPAAF 156

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	823	100.0	156	1 IEX1_HUMAN	P46695 homo sapien
2	525	63.8	153	1 IEX1_MOUSE	P46694 mus musculus
3	112.5	13.7	283	1 EXTN_SORBI	P24152 sorghum bic
4	105	12.8	3149	1 TEGU_EBV	P03186 epstein-bar
5	104	12.6	815	1 MK07_HUMAN	Q13164 homo sapien
6	102.5	12.5	381	1 MODD_MYCAV	Q48919 mycobacteri
7	102.5	12.5	3866	1 HRX_MOUSE	P55200 mus musculus
8	101.5	12.3	268	1 NO20_MEDTR	P93329 medicago tr
9	100	12.2	620	1 EXTN_TOBAC	P13983 nicotiana t
10	99.5	12.1	279	1 Y091_NPVAC	O10341 orgyia pseu
11	98.5	12.0	3164	1 TEGU_HSV11	P10220 herpes simp
12	95.5	11.6	537	1 MYPH_CHICK	Q05623 gallus gall
13	95	11.5	439	1 XP2_XENLA	P17437 xenopus lae
14	94.5	11.5	510	1 ERM_HUMAN	P41161 homo sapien
15	94.5	11.5	758	1 HUNB_DROME	P05084 drosophila
16	94.5	11.5	1595	1 SOS_DROME	P26675 drosophila
17	94	11.4	1290	1 PER1_HUMAN	O15534 homo sapien
18	93.5	11.4	1402	1 IF4G_RABIT	P41110 oryctolagus
19	93	11.3	817	1 VRP1_YEAST	P37370 saccharomyc
20	92.5	11.2	605	1 BR11_EBV	P03209 epstein-bar
21	92.5	11.2	625	1 NIFA_AZOBR	P30667 azospirillum
22	92	11.2	534	1 ARX_ARATH	P40602 arabidopsis
23	92	11.2	3969	1 HRX_HUMAN	Q03164 homo sapien
24	91.5	11.1	267	1 EXTN_MAIZE	P14918 zes mays (m
25	91.5	11.1	615	1 MUTL_ECOLI	P23367 escherichia
26	91.5	11.1	1133	1 SRE1_CRIGR	O60416 cricetus
27	90	10.9	316	1 CDNC_HUMAN	P49918 homo sapien
28	90	10.9	449	1 APC_BRANA	P40603 brassica na
29	89.5	10.9	1487	1 IC44_HSVK	P17473 equine herp
30	89.5	10.9	1715	1 TRX2_HUMAN	Q9um66 homo sapien
31	89	10.8	1291	1 PER1_MOUSE	O35973 mus musculus
32	88.5	10.8	865	1 CPN_DROME	Q02910 drosophila
33	88	10.7	474	1 VTP3_TTVIV	P19275 thermoprote

## ALIGNMENTS

RESULT 1  
ID IEX1\_HUMAN STANDARD: PRT: 156 AA.  
AC P46695; Q93044; Q92691;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY  
DE PROTEIN GLY96) (IMMEDIATE EARLY RESPONSE 3 PROTEIN) (PACAP-RESPONSIVE  
DE GENE 1 PROTEIN) (PRG1 PROTEIN) (DIFFERENTIATION-DEPENDENT GENE 2  
DE PROTEIN) (DIF-2 PROTEIN).  
GN IER3 OR IEX1 OR PRG1 OR DIF2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE=96181295; PubMed=8603392;  
RA Kondratyev A.D.; Chung K.-N.; Jung M.O.;  
RT Identification and characterization of a radiation-inducible  
RT glycosylated human early-response gene.";  
RL Cancer Res. 56:1498-1502(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96221139; PubMed=8653710;  
RA Schaefer H.; Trauzold A.; Siegel E.G.; Folsch U.R.; Schmidt W.E.;  
RT "PRG1: a novel early-response gene transcriptionally induced by  
RT pituitary adenylate cyclase activating polypeptide in a pancreatic  
RT carcinoma cell line".  
RL Cancer Res. 56:2641-2641(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97339426; PubMed=9196025;  
RA Pietzsch A.; Buechler C.; Aslanidis C.; Schmitz G.;  
RT Identification and characterization of a novel monocyte/macrophage  
RT differentiation-dependent gene that is responsive to  
RT lipopolysaccharide, ceramide, and lysophosphatidylcholine.";  
RL Biochem. Biophys. Res. Commun. 235:4-9(1997).  
RN [4]  
RP PRELIMINARY SEQUENCE OF 1-106 FROM N.A.  
RC TISSUE=PLACENTA;  
RA Hillier L.; Clark N.; Dubuque T.; Elliston K.; Hawkins M.; Holman M.;  
RA Hultman M.; Kucaba T.; Le M.; Lennon G.; Marra M.; Parsons J.;  
RA Rifkin L.; Rohlfing T.; Tan F.; Trevaskis E.; Waterston R.;  
RA Williamson A.; Wohldmann P.; Wilson R.;  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).  
CC -!- INDUCTION: BY RADIATION, TPA, OKADAIC ACID AND TNF-ALPHA.  
CC -!- PTM: GLYCOSYLATED.  
CC -!- SIMILARITY: STRONG, TO MOUSE ORTHOLOG.  
CC -!- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO  
CC FRAMESHIFTS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

## ALIGNMENTS

34 88 10.7 477 1 BIAR\_HUMAN P08588 homo sapien  
35 87.5 10.6 543 1 VP61\_NPVAC Q03209 autographa  
36 87.5 10.6 591 1 MNT1\_MOUSE O08789 mus musculus  
37 87 10.6 1575 1 SYJ1\_HUMAN O43426 homo sapien  
38 86.5 10.6 2414 1 P300\_HUMAN Q09472 homo sapien  
39 86.5 10.5 1023 1 SRE1\_RAT P56720 rattus norv  
40 86.5 10.5 1487 1 ICP4\_HSVEB P28925 equine herp  
41 86 10.4 390 1 VGLI\_HSV11 P06487 herpes simp  
42 86 10.4 1324 1 IRS2\_HUMAN O974h2 homo sapien  
43 85.5 10.4 224 1 Y091\_NPVAC P1479 autographa  
44 85.5 10.4 626 1 GPBA\_HUMAN P07359 homo sapien  
45 85.5 10.4 1233 1 NME3\_HUMAN Q14957 homo sapien





DR PROSITE: PS50014; BROMODOMAIN.2; 1.  
 KW Proto-oncogene: Chromosomal translocation; DNA-binding;  
 KW Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;  
 KW Alternative splicing; Polymorphism.

FT NON\_TER 1  
 FT DNA\_BIND 67 78 A.T HOOK (BY SIMILARITY).  
 FT DNA\_BIND 115 125 A.T HOOK (BY SIMILARITY).  
 FT DNA\_BIND 199 207 A.T HOOK (BY SIMILARITY).  
 FT DOMAIN 1330 1381 PHD 1.  
 FT DOMAIN 1383 1432 PHD 2.  
 FT DOMAIN 1465 1529 PHD 3.  
 FT DOMAIN 1605 1650 BROMODOMAIN (DIVERGENT).  
 FT DOMAIN 3737 3846 SET DOMAIN.  
 FT DOMAIN 35 41 POLY-GLY.  
 FT DOMAIN 459 469 PRO-RICH.  
 FT DOMAIN 1231 1238 POLY-PRO.  
 FT DOMAIN 3533 3536 POLY-GLU.  
 FT DOMAIN 3693 3697 POLY-GLU.  
 FT VARSPLIC 1503 1505 MISSING (IN ISOFORM 2).  
 FT VARIANT 1497 1497 K -> T.  
 SQ SEQUENCE 3866 AA; 420976 MW; ADFC55E14E806F1D CRC64;

Query Match 12.5%; Score 102.5; DB 1; Length 3866;

Best Local Similarity 26.8%; Pred. No. 1.8;

Matches 41; Conservative 19; Mismatches 48; Indels 45; Gaps 8;

QY 12 TILQAP-TPAPSTIPGRSGPEIFTDPLPE-----PAAAPAGRPS--ASR 56  
 DB 1129 TSVKSPLEPAQAARPPREPAPKSSSEPPRPKPKVEKSEGGAPAPAPPEPKQVSAP 1180  
 QY 57 GHRKSRRLYPRVVRQLPVEPNPAKRLLELLTIVFCQILMAEGVPA-----PLP 110  
 DB 1189 ASRKSSKQVSQPAV---VPPQPPSTAP-----OKKEAPKAVSEPKKKQPPP 1233  
 QY 111 PEDAPNAA---SLAPTPSPVLEPNLTSEPSD 140  
 DB 1234 PEPGPEQSKQKVAPLPSIPVKQ-----KPKD 1260

RESULT 8  
 ID NO20\_MEDTR STANDARD: PRT: 268 AA.  
 AC N9020\_MEDTR  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE EARLY NODULIN 20 PRECURSOR (N-20).  
 GN ENO20.  
 OS Medicago truncatula (Barrel medic).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 OC Fabales; Fabaceae; Papilionoideae; Medicago.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. JEMALONG;  
 RA MEDLINE=96187258; PubMed=9526510;  
 RA Greene E.A., Erard M., Dedieu A., Barker D.G.;  
 RT "Mcnod16 and 20 are members of a family of phytoecyanin-related early nodulins".  
 RL Plant Mol. Biol. 36:775-783(1998).  
 CC -!- SIMILARITY: CONTAINS ONE PLASTOCYANIN-LIKE DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: X99467; CAA67830.1; -  
 DR Nodulation; Signal.

FT SIGNAL 1 22 BY SIMILARITY.  
 FT CHAIN 23 268 EARLY NODULIN 20.  
 FT DOMAIN 123 145 PLASTOCYANIN-LIKE.  
 FT DOMAIN 136 145 POLY-PRO.  
 SQ SEQUENCE 268 AA; 28668 MW; 6FA29C5798C75F91 CRC64;  
 Query Match 12.3%; Score 101.5; DB 1; Length 268;  
 Best Local Similarity 28.3%; Pred. No. 0.15;  
 Matches 41; Conservative 12; Mismatches 31; Indels 61; Gaps 9;  
 QY 13 ILQAPTAPS-----TIPGRSGPEIFTDPLPEPAAAPAGRPSASRGRKRERR 64  
 DB 132 VLSSPPPPSPPTPRSSPTPIPHPPRSPLS---PPSPSPSPSPSPSPSPSPS 178  
 QY 65 VLYPRVVRQLPVEPNPAKRLLELLTIVFCQILMAEGVPAIP-----PEDAPNA 117  
 DB 179 -----PRSTPI--PHPRKR-----SPASPSPSPLSKSPSPSESP-- 211  
 QY 118 ASLAPTPVSPV--LEPFLNLTSEPSD 140  
 DB 212 -SLAPSPSDSVASLAP---SSSPSD 232

RESULT 9

EXTN\_TOBAC

ID EXTN\_TOBAC STANDARD: PRT: 620 AA.  
 AC P13983;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-MAR-1992 (Rel. 21, Last annotation update)  
 DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).  
 GN HRCPT3.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 OC Solanales; Solanaceae; Nicotiana.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. XANTHI; TISSUE=LEAF;  
 RA MEDLINE=90128263; PubMed=2612909;  
 RA Keller B., Lamb C.J.;  
 RT "Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene in lateral root initiation".  
 RL Genes Dev. 3:1639-1646(1989).  
 CC -!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE MAIN ROOT.  
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.  
 CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN GLYCOSYLATED.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: X13885; CAA32090.1; -  
 DR PIR: S06733; S06733.  
 KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;  
 KW Hydroxylation.  
 FT SIGNAL 1 ?  
 FT CHAIN ? 620 EXTENSIN.  
 FT REPEAT 70 73 H-A-P-P.  
 FT REPEAT 148 151 H-A-P-P.  
 FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.  
 FT REPEAT 229 235 1.  
 FT REPEAT 236 262 2.  
 FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.

RT Molecular cloning of chicken myosin-binding protein (MyBP) H (86-kDa  
RT protein) reveals extensive homology with MyBP-C (C-protein) with  
RL conserved immunoglobulin C2 and fibronectin type III motifs.";  
RL J. Biol. Chem. 268:3670-3676(1993).  
CC -!- FUNCTION: BINDS TO MYOSIN; PROBABLY INVOLVED IN INTERACTION WITH  
CC -!- THICK MYOFILAMENTS IN THE A-BAND.  
CC -!- TISSUE SPECIFICITY: SKELETAL MUSCLE. SEEMS TO BE ALSO EXPRESSED IN  
CC THE SLOW TONIC ALD MUSCLE. NOT DETECTED IN GIZZARD OR HEART.  
CC -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN-TYPE III-LIKE DOMAINS.  
CC -!- SIMILARITY: BELONGS TO THE MYBP FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L05605; AAA21418.1;  
DR INTERPRO: IPR001777;  
DR INTERPRO: IPR002965;  
DR INTERPRO: IPR003006;  
DR PFAM: PF00041; fn3; 2.  
DR PFAM: PF00047; 19; 2.  
DR PRINTS: PR00014; FNTYPEIII.  
DR PRINTS: PR01217; PRICHEXTENSN.  
KW Immunoglobulin domain; Cell adhesion; Muscle protein; Thick filament;  
KW Repeat.  
FT DOMAIN 135 221 FIBRONECTIN TYPE-III.  
FT DOMAIN 253 312 IG-LIKE C2-TYPE DOMAIN.  
FT DOMAIN 331 416 FIBRONECTIN TYPE-III.  
FT DOMAIN 458 518 IG-LIKE C2-TYPE DOMAIN.  
FT CONFLICT 2 2 T -> G (IN AA SEQUENCE).  
FT CONFLICT 9 9 A -> P (IN AA SEQUENCE).  
FT CONFLICT 15 15 A -> K (IN AA SEQUENCE).  
SQ SEQUENCE 537 AA; 58678 MW; 06C4CF0FE1DD233 CRC64;

Query Match 11.6%; Score 95.5; DB 1; Length 537;  
Best Local Similarity 27.5%; Pred. No. 0.89;  
Matches 38; Conservative 6; Mismatches 61; Indels 33; Gaps 5;  
QY 9 PTMTLQAPTPAPSTIPGRRGSGPEIFTDFPLPEPAAPAGPSASGRHKRSRRVLYP 68  
DB 38 PTPKSGHAPTPEKHAPPKEEHAP-----PPKECHAPAPAAETPPAHEHPDAPQPAAP 92  
QY 69 RVRRQLPVEEPNPAKRLFLLLTFVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPV 128  
DB 93 AAEHAPTTHAAPAH-----EEG-PPAPAPAP-----APEP----- 125  
QY 129 LEFNLTSRPSDYALDLS 146  
DB 126 -EPEKPEEPSVPLSLA 142

RESULT 13  
XP2\_XENLA STANDARD; PRT; 439 AA.  
AC P17437; Q08944;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE SKIN SECRETORY PROTEIN XP2 PRECURSOR (APEG PROTEIN).  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;  
OC Xenopodinae; Xenopus.  
RN [1]  
RP SEQUENCE OF 1-25 AND 344-439 FROM N.A.  
RX MEDLINE=92332564; PubMed=1629230;  
RA Hauser F., Roeben C., Hoffmann W.;

RT "xp2, a new member of the P-domain peptide family of potential growth  
RT factors, is synthesized in Xenopus laevis skin.";  
RL J. Biol. Chem. 267:14451-14455(1992).  
RN [2]  
RP SEQUENCE OF 3-439 FROM N.A.  
RT TISSUE-SKIN;  
RX MEDLINE=90127399; PubMed=22982993;  
RA Gnaehl M., Berger H., Thalhammer J., Kreil G.;  
RT "Dermal glands of Xenopus laevis contain a polypeptide with a highly  
RT repetitive amino acid sequence.";  
RL FEBS Lett. 260:145-148(1990).  
CC -!- FUNCTION: MAY ACT AS A GROWTH FACTOR IN THE GERMINAL LAYER OF THE  
CC EPIDERMIS. MAY ALSO BE INVOLVED IN GROWTH OF REGENERATING GLANDS  
CC AND IN PROTECTION OF THE SKIN FROM THE EXTERNAL ENVIRONMENT.  
CC -!- SUBCELLULAR LOCATION: SECRETED.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS OF THE PROTEINS ARE  
CC PRODUCED BY ALTERNATIVE SPLICING. THE APEG FORM INCLUDES A LARGE  
CC DOMAIN WITH 33 SEQUENCE REPEATS.  
CC -!- TISSUE SPECIFICITY: SKIN.  
CC -!- SIMILARITY: CONTAINS 2 P-TYPE (TREFOL) DOMAINS.  
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 392  
CC ONWARD AND IS SHORTER (418 AA) DUE TO A FRAMESHIFT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M90095; AAA50001.1;  
DR EMBL: X51394; CAA35759.1; ALT\_FRAME.  
DR PIR: S07498; SKXLG.  
DR PIR: A37331; A37331.  
DR HSP: P04155; IFS2.  
DR INTERPRO: IPR000519;  
DR PFAM: PF00088; trefol; 2.  
DR PRINTS: PR00680; PTREFOL.  
DR PROSITE: PS00025; P\_TREFOL; 2.  
KW Signal; Growth factor; Alternative splicing; Amphibian skin; Repeat.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 439 SKIN SECRETORY PROTEIN XP2.  
FT MOD\_RES 23 23 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).  
FT DOMAIN 26 343 33 X REPEATS OF G-[GE]-[AP](2,4)-A-E.  
FT DOMAIN 350 391 P-TYPE 1.  
FT DOMAIN 397 438 P-TYPE 2.  
FT DISULFID 351 377 BY SIMILARITY.  
FT DISULFID 361 376 BY SIMILARITY.  
FT DISULFID 371 388 BY SIMILARITY.  
FT DISULFID 398 424 BY SIMILARITY.  
FT DISULFID 408 423 BY SIMILARITY.  
FT DISULFID 418 435 BY SIMILARITY.  
FT VARSPIC 26 343 MISSING (IN ISOFORM XP2).  
FT CONFLICT 3 3 H -> S (IN REF. 2).  
FT CONFLICT 18 18 C -> W (IN REF. 2).  
SQ SEQUENCE 439 AA; 41173 MW; 38C4A4B57CBAE778 CRC64;

Query Match 11.5%; Score 95; DB 1; Length 439;  
Best Local Similarity 25.7%; Pred. No. 0.8;  
Matches 35; Conservative 6; Mismatches 57; Indels 38; Gaps 5;  
QY 15 QAPTAP-----STIPGRRGSGPEIFTDFPLPEPAAPAGPSASGRHKRSRRVLYPRV 70  
DB 247 EAPAPAEAGEAPAPAEAGEAPA-----PAPAEAGEAPAPAEAGEAPAEAGEAPAP 301  
QY 71 VRROLPE-----EPNPAKRLFLLLTFVFCQILMAEGVPAPLPEDAPNAASLAPTP 124  
DB 302 AEGEAPAPAEAGGAPSP-----AEGGAPAAAPAEAGGAPAPAPAE 343  
QY 125 VSPVLE-----PFNLT 135

Search completed: January 30, 2001, 23:12:24  
Job time: 4005 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2001, 23:12:24 ; Search time 43.61 Seconds  
(without alignments)  
63.685 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156

Perfect score: 442  
Sequence: 1 VRQLPVEPNPAKRLFL.....EPSDYALDSTLQHQHAAAF 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	442	100.0	156	1	1	1	P46695 homo sapien
2	273	61.8	153	1	1	1	P46694 mus musculus
3	78.5	17.8	1595	1	1	1	P26675 drosophila
4	73.5	16.6	272	1	1	1	P31698 erwinia chr
5	72.5	16.4	272	1	1	1	Q01564 erwinia chr
6	71.5	16.2	626	1	1	1	P07359 homo sapien
7	70.5	16.0	605	1	1	1	P03209 epstein-bar
8	70.5	16.0	2715	1	1	1	Q9umh6 homo sapien
9	69.5	15.7	1894	1	1	1	P19097 saccharomyc
10	69	15.6	1402	1	1	1	P41110 oryctolagus
11	68.5	15.5	181	1	1	1	Q9v426 homo sapien
12	68.5	15.5	882	1	1	1	P12830 homo sapien
13	68.5	15.5	1337	1	1	1	Q12913 homo sapien
14	68.5	15.5	1844	1	1	1	P10358 turnip yell
15	68	15.4	1395	1	1	1	Q04637 homo sapien
16	67	15.2	283	1	1	1	P24152 sorghum bic
17	67	15.2	586	1	1	1	P25579 saccharomyc
18	67	15.2	4289	1	1	1	P22105 homo sapien
19	66.5	15.0	3149	1	1	1	P03186 epstein-bar
20	66	14.9	621	1	1	1	Q59498 mycobacteri
21	65.5	14.8	325	1	1	1	Q30620 mycobacteri
22	65.5	14.8	325	1	1	1	Q05096 mycobacteri
23	65.5	14.8	329	1	1	1	Q05605 pseudomonas
24	65.5	14.8	812	1	1	1	P12977 epstein-bar
25	65.5	14.8	1093	1	1	1	P55198 homo sapien
26	65	14.7	209	1	1	1	Q15255 homo sapien
27	65	14.7	328	1	1	1	P71773 escherichia
28	65	14.7	427	1	1	1	Q15561 homo sapien
29	65	14.7	724	1	1	1	P23726 bos taurus
30	64.5	14.6	686	1	1	1	Q03237 gallus gall
31	64.5	14.6	1183	1	1	1	P54258 rattus norv
32	64	14.5	510	1	1	1	O81973 glycine max
33	64	14.5	510	1	1	1	Q01101 homo sapien

34 64 14.5 537 1 PTN5\_HUMAN  
35 64 14.5 541 1 PTN5\_MOUSE  
36 64 14.5 590 1 FTZ1\_RHIME  
37 64 14.5 966 1 SSN6\_YEAST  
38 63.5 14.4 443 1 HX43\_HUMAN  
39 63.5 14.4 540 1 NUSA\_MYCPN  
40 63.5 14.4 816 1 QALF\_NEUCR  
41 63.5 14.4 1776 1 POLR\_OYMW  
42 63.5 14.4 1844 1 POLR\_TYMW  
43 63 14.3 323 1 PF27\_MOUSE  
44 63 14.3 420 1 EFLA\_HALWA  
45 63 14.3 598 1 NR41\_HUMAN

## ALIGNMENTS

RESULT 1  
ID IEX1\_HUMAN STANDARD; PRT: 156 AA.  
AC P46695; Q93044; Q92691;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY  
DE PROTEIN GLY96) (IMMEDIATE EARLY RESPONSE 3 PROTEIN) (PACAP-RESPONSIVE  
DE GENE 1 PROTEIN) (PRG1 PROTEIN) (DIFFERENTIATION-DEPENDENT GENE 2  
DE PROTEIN) (DIF-2 PROTEIN).  
GN IER3 OR IEX1 OR PRG1 OR DIF2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RP [1]  
RC SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE=96181295; PubMed=8603392;  
RA Kondratyev A.D., Chung K.-N., Jung M.O.;  
RT "Identification and characterization of a radiation-inducible  
RT glycosylated human early-response gene.";  
Cancer Res. 56:1498-1502(1996).  
RL [2]  
RC SEQUENCE FROM N.A.  
RX MEDLINE=96221139; PubMed=8653710;  
RA Schaefer H., Trauzold A., Siegel E.G., Folsch U.R., Schmidt W.E.;  
RT "PRG1: a novel early-response gene transcriptionally induced by  
RT pituitary adenylate cyclase activating polypeptide in a pancreatic  
RT carcinoma cell line.";  
Cancer Res. 56:2641-2641(1996).  
RL [3]  
RC SEQUENCE FROM N.A.  
RX MEDLINE=97339426; PubMed=9196025;  
RA Pietzsch A., Buechler C., Aslanidis C., Schmitz G.;  
RT "Identification and characterization of a novel monocyte/macrophage  
RT differentiation-dependent gene that is responsive to  
RT lipopolysaccharide, ceramide, and lysophosphatidylcholine.";  
Biochem. Biophys. Res. Commun. 235:4-9(1997).  
RL [4]  
RC PRELIMINARY SEQUENCE OF 1-106 FROM N.A.  
RC TISSUE=PLACENTA;  
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman M.,  
RA Hultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J.,  
RA Rifkin L., Rohlfing T., Tan F., Trevaskis E., Waterston R.,  
RA Williamson A., Wohlmann P., Wilson R.;  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).  
CC -!- INDUCTION: BY RADIATION, TPA, OKADAIC ACID AND TNF-ALPHA.  
CC -!- SIMILARITY: STRONG, TO MOUSE ORTHOLOG.  
CC -!- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO  
FRAMESHIFTS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```
DR FLYBASE: FBgn0001965; Sos.
DR INTERPRO: IPR000219; -.
DR INTERPRO: IPR000651; -.
DR INTERPRO: IPR001849; -.
DR INTERPRO: IPR001895; -.
DR PFAM: PF00169; PH; 1.
DR PFAM: PF00617; RasGEF; 1.
DR PFAM: PF00621; RhogEF; 1.
DR PFAM: PF00720; GDS_CDC35; 1.
DR PROSITE: PS00003; PH_DOMAIN; 1.
DR KW Guanine-nucleotide releasing factor; Neurogenesis.
FT DOMAIN 15 51
FT DOMAIN 248 478 DH.
FT DOMAIN 479 587 PH.
FT DOMAIN 1511 1516 GLN-RICH.
FT DOMAIN 1525 1541 HIS-RICH.
FT CONFLICT 232 243 TSCPVCHFPDR -> HILPSLSLPAQR
FT CONFLICT 1462 1462 V -> P (IN REF. 2).
FT SEQUENCE 1595 AA; 177837 MW; 33AE31F0767A219F CRC64;

Query Match 17.88; Score 78.5; DB 1; Length 1595;
Best Local Similarity 28.38; Pred. No. 3.9;
Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps 6;

QY 1 VRRQLPVE-----EPNPAKRLFLLLTIVFCOILMNEGVPAPLPPEDAPNAAASLAPT 53
DB 1472 IRRNSAIKRAAATSOPIAAGISTTLTVSOAVATDEVLPPLISP-----AASSSTT 1476

QY 54 --PVSPLVEF--NLTSPE-----SDYALDLSTFLQO-----HPAAF 86
DB 1477 TSPLTPAMSPNIPSHPVSTSSSYAHQLRMRQOQQOQTHPAIY 1522

RESULT 4
GSPC_ERWCH STANDARD; PRT; 272 AA.
AC P31698;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE GENERAL SECRETION PATHWAY PROTEIN C (PECTIC ENZYMES SECRETION PROTEIN
DE OUTC).
GN OUTC.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EC16;
RX MEDLINE=93054355; PubMed=1429461;
RA "Analysis of eight out genes in a cluster required for pectic enzyme
RT secretion by Erwinia chrysanthemi: sequence comparison with secretion
RT genes from other gram-negative bacteria.";
RL J. Bacteriol. 174:7385-7397(1992).
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF THE
CC MULTIPLE PECTIC ENZYMES.
CC -!- SUBCELLULAR LOCATION: INNER MEMBRANE (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE EXEC/PULC/OUTC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L02214; AAA24830.1; .
DR EMBL; L02214; AAA24830.1; .

Query Match 16.68; Score 73.5; DB 1; Length 272;
Best Local Similarity 30.18; Pred. No. 1.7;
Matches 22; Conservative 13; Mismatches 13; Indels 13; Gaps 3;

QY 4 QLPVEEPNPAKRLFLLLTIVFCQ---ILMAEAGVPAPLPPEDAPNAAASLAPTVPSPVLE 60
DB 5 KLPPLSPSVIRILFVLLMLLFCQLAMIFMRVGLP-----DNSPVASVQITPAOAROQ 58

QY 61 PFNLTSEPSDYAL 73
DB 59 PVTLL-----NDFTL 67

RESULT 5
GSPC_ERWCH STANDARD; PRT; 272 AA.
AC Q01584;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE GENERAL SECRETION PATHWAY PROTEIN C (PECTIC ENZYMES SECRETION PROTEIN
DE OUTC).
GN OUTC.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5937;
RX MEDLINE=93086427; PubMed=1453958;
RA Condelline G.; Dorel C.; Hugouvioux-Cotte-Pattat N.; Robert-Baudouy J.;
RT "Some of the out genes involved in the secretion of pectate lyases in
RT Erwinia chrysanthemi are regulated by kdgr.";
RL Mol. Microbiol. 6:3199-3211(1992).
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF THE
CC MULTIPLE PECTIC ENZYMES.
CC -!- SUBCELLULAR LOCATION: INNER MEMBRANE (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE EXEC/PULC/OUTC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X65265; CAA46369.1; .
DR EMBL; X65265; CAA46369.1; .
DR PIR; S28013; S28013.
DR INTERPRO: IPR001478; .
DR INTERPRO: IPR001639; .
DR PFAM: PF00595; PD2; 1.
DR PRINTS: PR00810; BCTERIALGSPC.
DR PROSITE: PS01141; T2SP-C; 1.
DR TRANSMEM 1 16 CYTOPLASMIC (POTENTIAL).
KW Transport; Transmembrane; Inner membrane.
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 35 PERIPLASMIC (POTENTIAL).
FT DOMAIN 36 272 PERIPLASMIC (POTENTIAL).
FT SEQUENCE 272 AA; 30301 MW; 064C7311F9714405 CRC64;
```

ACTIN-BINDING PROTEIN.  
 -1- MISCELLANEOUS: BINDING SITES FOR VON WILLEBRAND FACTOR AND THROMBIN (THE LATTER SITE WITH UNKNOWN FUNCTION) ARE IN THE AMINO-TERMINAL PART OF THE MOLECULE.  
 -1- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN MANY PROTEINS. NUMBER IN THIS PROTEIN: 8.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: J02940; AA52595.1; -  
 EMBL: M22403; AA52596.1; -  
 PIR: A27075; NBUHA.  
 MIM: 231200; -  
 MIM: 177820; -  
 DR INTERPRO: IPR000372; -  
 DR INTERPRO: IPR000483; -  
 DR INTERPRO: IPR001611; -  
 DR PFAM: PF00560; LRR; 6.  
 DR PFAM: PF01463; LRRCT; 1.  
 DR PFAM: PF01462; LRRNT; 1.  
 DR PRINTS: PR00019; LEURICHRPT.  
 KW Platelet; Transmembrane; Glycoprotein; Hemostasis; Blood coagulation;  
 KW Repeat; Leucine-repeat; Signal; Cell adhesion; Disease mutation;  
 KW Polymorphism; von Willebrand disease; Bernard Soulier syndrome.  
 FT SIGNAL 1 16  
 FT CHAIN 1 626 PLATELET GLYCOPROTEIN IB ALPHA CHAIN.  
 FT CHAIN 17 626 GLYCOCALICIN.  
 FT DOMAIN 17 505 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 506 526 POTENTIAL.  
 FT DOMAIN 527 626 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 39 224 LEUCINE-RICH REPEATS.  
 FT REPEAT 39 60 LRR 1.  
 FT REPEAT 61 83 LRR 2.  
 FT REPEAT 84 108 LRR 3.  
 FT REPEAT 109 131 LRR 4.  
 FT REPEAT 132 155 LRR 5.  
 FT REPEAT 156 179 LRR 6.  
 FT REPEAT 180 204 LRR 7.  
 FT REPEAT 205 224 LRR 8.  
 FT REPEAT 379 386 THR/PRO-RICH.  
 FT REPEAT 387 395 THR/PRO-RICH.  
 FT REPEAT 400 408 THR/PRO-RICH.  
 FT REPEAT 409 417 THR/PRO-RICH.  
 FT REPEAT 422 430 THR/PRO-RICH.  
 FT DISULFID 20 33  
 FT DISULFID 225 264  
 FT DISULFID 227 280  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 308 308  
 FT VARIANT 73 73  
 FT VARIANT 161 161 /FTID-VAR\_005256.  
 FT VARIANT 172 172 T -> M (IN ALLOANTIGEN SIBA(+)).  
 FT VARIANT 172 172 A -> V (IN BSS).  
 FT VARIANT 195 195 /FTID-VAR\_005258.  
 FT VARIANT 225 225 MISSING (IN BSS).  
 FT VARIANT 225 225 /FTID-VAR\_005259.  
 FT VARIANT 249 249 C -> S (IN BSS).  
 FT VARIANT 249 249 /FTID-VAR\_005260.  
 FT VARIANT 249 249 G -> V (IN PSEUDO-VWD).  
 FT VARIANT 255 255 /FTID-VAR\_005261.  
 FT VARIANT 255 255 M -> V (IN PSEUDO-VWD).  
 FT VARIANT 255 255 /FTID-VAR\_005262.  
 SQ SEQUENCE 626 AA; 68955 MW; 19514119845DF573 CRC64;  
 Query Match 16.2%; Score 71.5; DB 1; Length 626;

Best Local Similarity 45.0%; Pred. No. 6.8;  
 Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;  
 QY 36 PAPLP-----PDAPNAASLAPTPVSPVLEPFPNLTSEPS 69  
 DB 382 PTPSPATSEVPPEPAPNMTLTPTSPPTPEP---TSEPA 418  
 -----  
 RESULT 7  
 BRLL\_EBV  
 ID BRLL\_EBV STANDARD; PRT; 605 AA.  
 AC P03209;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE TRANSCRIPTION ACTIVATOR BRLL1.  
 GN BRLL1.  
 OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Lymphocryptovirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84270667; PubMed=6087149;  
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,  
 RA Tuffnell P.S., Barrell B.G.;  
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";  
 RL Nature 310:207-211(1984).  
 CC -1- FUNCTION: TRANSCRIPTION ACTIVATION.  
 CC -1- MISCELLANEOUS: BRLL1 PROTEIN IS AN EARLY OR IMMEDIATE EARLY  
 CC PROTEIN.  
 CC -1- SIMILARITY: TO HVS-1 EDRF1 (GENE 50).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 EMBL: V01555; CAA24814.1; -  
 DR PIR: A03771; Q0BE29.  
 DR PIR: S33019; S33019.  
 DR TRANSFAC: T00710; -  
 KW Transcription regulation; Activator; DNA-binding; Early protein.  
 SQ SEQUENCE 605 AA; 56594 MW; B587DF0593FAD808 CRC64;  
 Query Match 16.0%; Score 70.5; DB 1; Length 605;  
 Best Local Similarity 47.5%; Pred. No. 8.2;  
 Matches 19; Conservative 2; Mismatches 18; Indels 1; Gaps 1;  
 QY 36 PAPLPEDAPNAASLAPTPVSPVLEPFP-NLTSPSPDYALD 74  
 DB 453 PGSPWANRPLPASLAPTGTGPVHPVGLTPAPVQPLD 492  
 -----  
 RESULT 8  
 TRX2\_HUMAN  
 ID TRX2\_HUMAN STANDARD; PRT; 2715 AA.  
 AC Q9UMN6; Q9UK25; Q95836; Q9Y669; Q9Y668; O15022;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE TRITHORAX HOMOLOG 2 (MIXED LINEAGE LEUKEMIA GENE HOMOLOG 2 PROTEIN)  
 DE (K1AA0340).  
 GN TRX2 OR HRX2 OR MLL2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A. (LONG ISOFORM).

```
RT "Mapping the cleavage site in protein synthesis initiation factor  
RT eIF-4 gamma of the 2A proteases from human Cocksackievirus and  
RL rhinovirus.";  
RN J. Biol. Chem. 268:19200-19203(1993).  
[2]  
RN  
RP PARTIAL SEQUENCE.  
RA MEDLINE=33034654; Pubmed=1429670;  
RX Yan R., Rychlik W., Echolson D., Rhoads R.E.;  
RT "Amino acid sequence of the human protein synthesis initiation factor  
RT eIF-4 gamma.";  
RN J. Biol. Chem. 267:23226-23231(1992).  
CC -!- FUNCTION: COMPONENT OF THE PROTEIN COMPLEX EIF-4, WHICH IS  
CC INVOLVED IN THE RECOGNITION OF THE MRNA CAP, ATP-DEPENDENT  
CC UNWINDING OF 5'-TERMINAL SECONDARY STRUCTURE AND RECRUITMENT OF  
CC MRNA TO THE RIBOSOME.  
CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
CC -!- SIMILARITY: THE C-TERMINAL REGION IS SIMILAR TO THE N-TERMINAL  
CC REGION OF WHEAT EUKARYOTIC INITIATION FACTOR (ISO)4F SUBUNIT P02.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
DR ENBL: L22090; AAA31242.1; --  
DR INTERPRO: IPR000504; --  
DR PFAM: PF02020; IF5_eIF4_eif2; 1.  
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.  
KW Initiation factor; Protein biosynthesis; Phosphorylation;  
KW RNA-binding.  
FT DOMAIN 188 192 POLY-PRO.  
FT DOMAIN 262 275 POLY-GLU.  
FT DOMAIN 502 607 RNA-BINDING (RNP2) (BY SIMILARITY).  
FT DOMAIN 698 706 RNA-BINDING (RNP1) (BY SIMILARITY).  
FT DOMAIN 1393 1398 POLY-GLU.  
SQ SEQUENCE 1402 AA; 154050 MW; 7FD85D7E30519230 CRC64;
```

```
Query Match      15.6% Score 69; DB 1; Length 1402;
Best Local Similarity 30.2%; Pred.No. 29;
Matches 26; Conservative 7; Mismatches 21; Indels 32; Gaps 4;

Oy  4 QLPVEERNPAKRLLFLTLTIIVFCOILMAECVAPLPPEPDAPNAAASIAPTVVS-PVLPEPF 62
    |:||||| |         ||||| |
Db   100 QMPVEEPAPISR-----EAGEPYCLSP-----PTPLAEPILEVE 134

Oy  63 NLTSEP-----SDYALDLSTFLQOH 82
    |:| :| :| :| :| :| :|
Db   135 VTLSKPVPVFSESSSPQVLTPLASH 160

RESULT 11
C210_HUMAN
ID C210_HUMAN STANDARD; PRT; 181 AA.
AC Q9Y426;
DT 30-MAY-2000 (Rel. 39, Created)
DI 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROTEIN C210RF25 (FRAGMENT).
CN C210RF25;
CS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
[1]
RN RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Koehrer K., Beyer A., Mexes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (May-1999) to the EMBL/GenBank/DDIJ databases.
CC This SWISS-PROT entry is copyright It is produced through a collaboration
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 12:11:54 ; Search time 1900.13 Seconds  
(without alignments)  
3307.456 Million cell updates/sec

Title: us-08-799-910-9  
Perfect score: 1228  
Sequence: 1 ATGTGCTACTCTCGACGCTG.....AAAAAATACTCGAG 1228

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_bal: \*  
2: gb\_baz: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_ph: \*  
6: gb\_pl1: \*  
7: gb\_pl2: \*  
8: gb\_pl3: \*  
9: gb\_pr2: \*  
10: gb\_pr3: \*  
11: gb\_ro: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: em\_fun: \*  
15: em\_hum1: \*  
16: em\_hum2: \*  
17: em\_in: \*  
18: em\_om: \*  
19: em\_or: \*  
20: em\_ov: \*  
21: em\_pat: \*  
22: em\_ph: \*  
23: em\_pi: \*  
24: em\_ro: \*  
25: em\_sts: \*  
26: em\_sy: \*  
27: em\_un: \*  
28: em\_vi: \*  
29: gb\_ba3: \*  
30: gb\_in1: \*  
31: gb\_in2: \*  
32: gb\_in3: \*  
33: gb\_pi3: \*  
34: gb\_pr4: \*  
35: em\_bal: \*  
36: em\_ba2: \*  
37: em\_htg1: \*  
38: em\_htg2: \*  
39: em\_htg3: \*  
40: em\_htg4: \*  
41: em\_htg5: \*  
42: em\_htg6: \*  
43: em\_htg7: \*

44: em\_htg8: \*  
45: em\_htg9: \*  
46: em\_htg10: \*  
47: em\_hum3: \*  
48: em\_hum4: \*  
49: em\_hum5: \*  
50: em\_hum6: \*  
51: gb\_pr5: \*  
52: gb\_pr6: \*  
53: gb\_pr7: \*  
54: gb\_htg1: \*  
55: gb\_htg2: \*  
56: gb\_htg3: \*  
57: gb\_htg4: \*  
58: gb\_htg5: \*  
59: gb\_htg6: \*  
60: gb\_htg7: \*  
61: gb\_htg8: \*  
62: gb\_htg9: \*  
63: gb\_htg10: \*  
64: gb\_htg11: \*  
65: gb\_htg12: \*  
66: gb\_htg13: \*  
67: gb\_htg14: \*  
68: gb\_htg15: \*  
69: gb\_htg16: \*  
70: gb\_htg17: \*  
71: gb\_htg18: \*  
72: gb\_htg19: \*  
73: gb\_htg20: \*  
74: gb\_htg21: \*  
75: gb\_htg22: \*  
76: gb\_htg23: \*  
77: gb\_sts1: \*  
78: gb\_sts2: \*  
79: gb\_v11: \*  
80: gb\_v12: \*  
81: gb\_pat1: \*  
82: gb\_pat2: \*  
83: em\_htg0: \*  
84: gb\_htg24: \*  
85: gb\_pr8: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result NO.	Score	Query Match	Length	DB	ID	Description
1	1207.8	98.4	1230	53	HSDIP2	Y14551 Homo sapien
2	1152.2	93.8	1223	85	S81914	S81914 IEX-1-radia
3	1080.4	88.0	44118	9	AC006165	AC006165 Homo sapi
4	1080.4	88.0	90244	8	AB023051	AB023051 Homo sapi
5	1080.4	88.0	200000	51	AF000512	AF000512 Homo sapi
6	1048.6	85.4	1309	10	AF039067	AF039067 Homo sapi
7	1048.6	85.4	1693	10	AF071596	AF071596 Homo sapi
8	979.8	79.8	1864	53	HSPRG1	X96438 H.sapiens p
9	471	38.4	477	10	AF083421	AF083421 Homo sapi
10	298.8	24.3	343	52	HSA227914	AJ227914 Homo sapi
11	242.4	19.7	1758	11	RNPRG1	X96437 R.norvegicu
12	210.6	17.1	203476	70	AC074150	AC074150 Mus muscu
13	209	17.0	1938	11	MMGLY96	X67644 M.musculus
14	202	16.4	187868	62	AC022301	AC022301 Mus muscu
15	151	12.3	297	81	A75448	A75448 Sequence 11
16	151	12.3	297	81	A78427	A78427 Sequence 11
17	122	9.9	122	78	G82116	G82116 1899 Human
18	74	6.0	205512	74	AL355388	AL355388 Homo sapi
19	70.4	5.7	267	81	A75468	A75468 Sequence 11
20	70.4	5.7	267	81	A78447	A78447 Sequence 11
21	64	5.2	83276	65	AC026631	AC026631 Homo sapi

Db 372 GCGCCACCCCTGTGTCCCGCTCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGAC 431  
Qy 421 TAGCTCTGGACCTGAGCACTTCTCTCCAGCAACACCCCGCGCTTCTAATCTGTGACTC 480  
Db 432 TAGCTCTGGACCTGAGCACTTCTCTCCAGCAACACCCCGCGCTTCTAATCTGTGACTC 491  
Qy 481 CCGCGACTCCCCAANAAGATCCGAAAAACCAAGAAACACACAGCGGTACTGTGGG 540  
Db 492 CCGCGACTCCCCAANAAGATCCGAAAAACCAAGAAACACACAGCGGTACTGTGGG 551  
Qy 541 CGAGAGCGTATCCCAACTGGGACTTCCGAGGCACTTCAACTCAGAACACTACAGCGA 600  
Db 552 CGAGAGCGTATCCCAACTGGGACTTCCGAGGCACTTCAACTCAGAACACTACAGCGA 611  
Qy 601 GAGCCACCCGCTGCTGTAGGCGGAGCCAGGCGCACAGAGCCAGCGGCATAGAGCC 660  
Db 612 GAGCCACCCGCTGCTGTAGGCGGAGCCAGGCGCACAGAGCCAGCGGCATAGAGCC 671  
Qy 661 GAGGCACAGCCAGCTGGGGCTAGGCGGCTAGGCGGAGGAGAGCGTCTTAATTTATCT 720  
Db 672 GAGGCACAGCCAGCTGGGGCTAGGCGGCTAGGCGGAGGAGAGCGTCTTAATTTATCT 731  
Qy 721 TATGCTCTTAATTAATATTAATATTAATTAATTAATTAATTAATTAATTAATTTCT 780  
Db 732 TATGCTCTTAATTAATATTAATTAATTAATTAATTAATTAATTAATTAATTTCT 791  
Qy 781 GTACGTAATATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTCT 840  
Db 792 GTACGTAATATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTCT 851  
Qy 841 AGTCTCTTGGTATTTATGAGCTTTGTGGAGCTGTGGAAGCAGGACCTGGAAGCTGC 900  
Db 852 AGTCTCTTGGTATTTATGAGCTTTGTGGAGCTGTGGAAGCAGGACCTGGAAGCTGC 911  
Qy 901 GGCAGAGTACGACAGAAATGGGAGGACTCGGGTGGGAGGAGCCTCGCGGTGGGATG 960  
Db 912 GGCAGAGTACGACAGAAATGGGAGGACTCGGGTGGGAGGAGCCTCGCGGTGGGATG 971  
Qy 961 AAGTCTGCTGGTGGGCTGAAGTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCCG 1020  
Db 972 AAGTCTGCTGGTGGGCTGAAGTTAGGAGGTGACTGCATCTCCAGCATCTCAACTCCG 1031  
Qy 1021 TCTGCTACTGTGAGACTTCGGCGGACCATTAGGAGTACGATCGGTGAGATCTCTTCCA 1080  
Db 1032 TCTGCTACTGTGAGACTTCGGCGGACCATTAGGAGTACGATCGGTGAGATCTCTTCCA 1091  
Qy 1081 TCTTCTTGAAGTGGCTTTAGGTTAGGTTAGGTTAGGTTAGGTTAGGTTAGGTTAG 1140  
Db 1092 TCTTCTTGAAGTGGCTTTAGGTTAGGTTAGGTTAGGTTAGGTTAGGTTAGGTTAG 1151  
Qy 1141 CACGAGCGACTGTGAGATCGGCTAGTATGTTCTGTGAACACAAATAAATTCATTAC 1200  
Db 1152 CACGAGCGACTGTGAGATCGGCTAGTATGTTCTGTGAACACAAATAAATTCATTAC 1211  
Qy 1201 TGTCAAAAAAATAAATAA 1219  
Db 1212 TGTCTGCTAAAAAATAAATA 1230

RESULT 2  
S81914 1223 bp mRNA PRI 12-AUG-1996  
LOCUS IEX-1-radiation-inducible immediate-early gene [human, placenta,  
mRNA partial, 1223 nt].  
ACCESSION S81914  
VERSION S81914.1 GI:1488384  
KEYWORDS human placenta.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1223)  
AUTHORS Kondratyev,A.D., Chung,K.N. and Jung,M.O.

TITLE Identification and characterization of a radiation-inducible  
glycosylated human early-response gene  
JOURNAL Cancer Res. 56 (7), 1498-1502 (1996)  
MEDLINE 96181295  
REMARK GenBank staff at the National Library of Medicine created this  
entry [NCBI gisbq 177109] from the original journal article.  
This sequence comes from Fig. 2A.  
FEATURES  
source 1..1223  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
gene 19..489  
/gene="IEX-1"  
/note="radiation-inducible immediate-early gene"  
CDS 19..489  
/gene="IEX-1"  
/note="radiation-inducible immediate-early gene; This  
sequence comes from Fig. 2A"  
/codon\_start=1  
/protein\_id="AAB36278.1"  
/db\_xref="GI:1488385"  
/translation="MCHSRSCPTMTILOAITPAPSTIPVPRRSGPEITFDNLPPEP  
AAAPAGRPSGSRGHRKSRVLYPRVRRPLVPRPNPAKRLFLLLTVFCOLLMAE  
EGVRAPLPEDAPNAASLAPTPVSPVLEPNLTSEPSYALDLSTFLOOHPAF"  
BASE COUNT 249 a 368 c 342 g 264 t  
ORIGIN  
Query Match 93.8%; Score 1152.2; D8 B5; Length 1223;  
Best Local Similarity 99.4%; Pred. No. 3.7e-255;  
Matches 1198; Conservative 0; Mismatches 3; Indels 4; Gaps 4;  
Qy 1 ATGTGTCTCTCGAGCTGCCACCGGACCATGACCATCTCGAGGCCCGGACCCCGGCC 60  
Db 19 ATGTGTCTCTCGAGCTGCCACCGGACCATGACCATCTCGAGGCCCGGACCCCGGCC 78  
Qy 61 CACTCCACCATCCCGGAGCCCGGCGGCTCGGTCTGAGATCTTCACCTTCGACCT 120  
Db 79 CACTCCACCATCCCGGAGCCCGGCGGCTCGGTCTGAGATCTTCACCTTCGACCT 138  
Qy 121 CTCCCGGAGCCGACGCGCCCTCGCGGCGCCCGAGCCCTCTCGCGGCGCACCAAG 180  
Db 139 CTCCCGGAGCCGACGCGCCCTCGCGGCGCCCGAGCGGCTCTCGCGGCGCACCAAG 198  
Qy 181 CGCAGCGCAGGCTTCTTACCTTCAGTGGTTCGGCGGCGAGTGGCAGTTCGAGAACCG 240  
Db 199 CGCAGCGCAGGCTTCTTACCTTCAGTGGTTCGGCGGCGAGTGGCAGTTCGAGAACCG 258  
Qy 241 AACCAGCCAAAGGCTTCTTCTGCTCTCACCATCTCTTCTGCGGAGATCTTGATG 300  
Db 259 AACCAGCCAAAGGCTTCTTCTGCTCTCACCATCTCTTCTGCGGAGATCTTGATG 318  
Qy 301 GCTGAAGAGGCTGCGCGGCGCCCTCGCTCCAGAGGAGCCCTAACGCGCATCCCTG 360  
Db 319 GCTGAAGAGGCTGCGCGGCGCCCTCGCTCCAGAGGAGCCCTAACGCGCATCCCTG 378  
Qy 361 GCGCCACCCCTGTGTCCCGCTCTCGAGCCCTTTAATCTGACTTCGAGCCCTCGGAC 420  
Db 379 GCGCCACCCCTGTGTCCCGCTCTCGAGCCCTTTAATCTGACTTCGAGCCCTCGGAC 438  
Qy 421 TAGCTCTGAGACTTCTGAGCACTTCTCTCCAGCAACACCCCGCGCTTCTAATCTGACTC 480  
Db 439 TAGCTCTGAGACTTCTGAGCACTTCTCTCCAGCAACACCCCGCGCTTCTAATCTGACTC 498  
Qy 481 CCGCGACTCCCCAANAAGATCCGAAAAACCAAGAAACACACAGCGGTACTGTGGG 540  
Db 499 CCGCGACTCCCCAANAAGATCCGAAAAACCAAGAAACACACAGCGGTACTGTGGG 558  
Qy 541 CGAGAGCGTATCCCAACTGGGACTTCCGAGGCACTTGAACCTCAGAACACTACACCGA 600  
Db 559 CGAGAGCGTATCCCAACTGGGACTTCCGAGGCACTTGAACCTCAGAACACTACACCGA 618  
Qy 601 GAGCCACCCGCTGTGTAGGCGGAGCCGAGGCGCACAGAGCCGAGGCGCATAGAGACC 660

repeat_region	3941..4070	/rpt_family="Alu"	
repeat_region	complement(4449..4724)	/rpt_family="Alu"	
repeat_region	complement(5117..5389)	/rpt_family="Alu"	
repeat_region	7668..7913	/rpt_family="Alu"	
repeat_region	complement(7924..8226)	/rpt_family="Alu"	
repeat_region	8250..8551	/rpt_family="Alu"	
repeat_region	complement(8893..8973)	/rpt_family="Alu"	
repeat_region	complement(11829..12096)	/rpt_family="MIR"	
repeat_region	12416..12740	/rpt_family="Alu"	
repeat_region	complement(13181..13247)	/rpt_family="Alu"	
repeat_region	complement(13498..13647)	/rpt_family="L1"	
repeat_region	13869..14279	/rpt_family="Alu"	
repeat_region	18033..18284	/rpt_family="Alu"	
repeat_region	19178..19451	/rpt_family="Alu"	
repeat_region	19458..19749	/rpt_family="Alu"	
repeat_region	20862..21002	/rpt_family="Alu"	
repeat_region	21566..21832	/rpt_family="MERS"	
repeat_region	22231..22521	/rpt_family="Alu"	
repeat_region	complement(22616..22679)	/rpt_family="Alu"	
repeat_region	complement(22869..23145)	/rpt_family="Alu"	
repeat_region	complement(23478..23772)	/rpt_family="Alu"	
repeat_region	24184..24372	/rpt_family="Alu"	
repeat_region	24529..24813	/rpt_family="Alu"	
repeat_region	complement(25060..25342)	/rpt_family="Alu"	
repeat_region	complement(25367..25653)	/rpt_family="Alu"	
repeat_region	25787..26076	/rpt_family="Alu"	
repeat_region	26491..26762	/rpt_family="Alu"	
repeat_region	26866..27143	/rpt_family="Alu"	
repeat_region	complement(27620..27919)	/rpt_family="Alu"	
repeat_region	29525..29719	/rpt_family="MIR"	
repeat_region	29762..30333	/rpt_family="Alu"	
repeat_region	30365..30653	/rpt_family="Alu"	
repeat_region	31430..31722	/rpt_family="Alu"	
repeat_region	31765..31857	/rpt_family="Alu"	
repeat_region	32265..32563	/rpt_family="Alu"	
repeat_region	complement(32710..32858)	/rpt_family="Alu"	
repeat_region	complement(33374..33584)	/rpt_family="Alu"	

repeat_region	/rpt_family="MER2"		
repeat_region	complement(34547..34595)	/rpt_family="MIR"	
repeat_region	complement(36673..36954)	/rpt_family="Alu"	
repeat_region	complement(37002..37230)	/rpt_family="Alu"	
repeat_region	complement(40018..40170)	/rpt_family="Alu"	
variation	40452	/note="high quality variation versus 3' overlapping clone"	
repeat_region	40509..41109	/rpt_family="Alu"	
repeat_region	42609..42892	/rpt_family="Alu"	
repeat_region	43240..43370	/rpt_family="MIR"	
repeat_region	43493..43779	/rpt_family="Alu"	
BASE COUNT	11505 a 10560 c 11226 g 10827 t		
ORIGIN			
Query Match	88.0%; Score 1080.4; DB 9; Length 44118;		
Best Local Similarity	91.4%; Pred. No. 1e-238;		
Matches 1203; Conservative	0; Mismatches 1; Indels 112; Gaps 1;		
Qy	1	ATGTGTCACTCTCGACGTGCGACCCGACCATGACCATCTCTGAGATCTTCACCTTCGACCT	60
Db	14601	ATGTGTCACTCTCGACGTGCGACCCGACCATGACCATCTCTGAGAGCCGCGACCCGCGCC	14660
Qy	61	CCCTCCACCATCCGCGGACCCGCGGCGGTCCGGTCTGAGATCTTCACCTTCGACCT	120
Db	14661	CCCTCCACCATCCGCGGACCCGCGGCGGTCCGGTCTGAGATCTTCACCTTCGACCT	14720
Qy	121	CTCCGAGCCGCGACGCGCCCTTCGCGGCGCCGCGGCTCTCGCGGCGACCCGAAAG	180
Db	14721	CTCCGAGCCGCGACGCGCCCTTCGCGGCGCCGCGGCTCTCGCGGCGACCCGAAAG	14780
Qy	181	CGCAGCGCGAGGTTCTTACCTCGAGT-----	209
Db	14781	CGCAGCGCGAGGTTCTTACCTCGAGTGTGAGTATCGCGGAGTGGGCAITTCGCGGG	14840
Qy	210	-----	209
Db	14841	GTGGGTGCGCTGGAGTCACTGGGGAAGACCCGACCTCCAGAGGCTTCGACCTGACCTGT	14900
Qy	210	-----GGTCCGCGCGCCAGCTGCCACTCGAGGAACCGAACCCAGCC	248
Db	14901	CTCCTGTTTGTCTCTCCCTTAGGTCCGCGCCAGCTGCCAGTCGAGGAACCGAACCCAGC	14960
Qy	249	CAAAAGGCTTCTCTTTCTGTGCTCAGCATCGTCTTCTGCCAGATCTCTGATGGCTGAAGA	308
Db	14961	CAAAAGGCTTCTCTTTCTGTGCTCAGCATCGTCTTCTGCCAGATCTCTGATGGCTGAAGA	15020
Qy	309	GGGTGTCGCGCGCCCTGCTCCAGAGAGCGCCCTAACCGCCGATCCCTTCGGCGCCAC	368
Db	15021	GGGTGTCGCGCGCCCTGCTCCAGAGAGCGCCCTAACCGCCGATCCCTTCGGCGCCAC	15080
Qy	369	CCCTGTGTCCCGCTCTCGAGGCTTTAATCTGACTTCGGAGGCTTCGACTACGCT	428
Db	15081	CCCTGTGTCCCGCTCTCGAGGCTTTAATCTGACTTCGGAGGCTTCGACTACGCT	15140
Qy	429	GGACCTCAGCACTTTCTCTCCAGCAACCCGCGGCTTTCTTAATCTGACTTCGCGCACT	488
Db	15141	GGACCTCAGCACTTTCTCTCCAGCAACCCGCGGCTTTCTTAATCTGACTTCGCGCACT	15200
Qy	489	CCCCAAAAGATCCGAAAACCAAGAAACACCGGCTTCTGCTGGCGGAGCGG	548
Db	15201	CCCCAAAAGATCCGAAAACCAAGAAACACCGGCTTCTGCTGGCGGAGCGG	15260
Qy	549	TATCCCCAAGTGGGACTTCCGAGGCAACTTGAACCTCAGAACCTACAGCGGAGAGCGG	608











QY 312 TGTGCGGCGCCCTGCTCCAGAGAGAGCCCTTACGCGCGATCC-----T 359  
Db 1132 TGTGTCGACGCGCTGGCTCCGAGGATGCTACGAGCGCGGTGACACCTGAGCCCATTTTC 1191  
QY 360 GCGCGCCACCCCTGTGTCGCCCTCTCTCAGCGCTTTATCTGACTTCGGAGCCCTCGGA 419  
Db 1192 TGGGCCATTACTGCGCCCGCTCTCGAGCTTTGACCTTGACTCGGAGTCTCGGA 1251  
QY 420 CTACGCTTGGACCTGACGACTTCTCTCAGCAACACCGCGCGCTTCTAACTGTGACT 479  
Db 1252 CTATGCGCTGGATCTTAAAGCTTTCTTAGCAACATCGCGCGCTTCTAAACGCGATG 1311  
QY 480 CCGCGACTCCCAAAAGATCCGAAACACCAAGAACACACGCGCTGCTGCTGC 539  
Db 1312 GGTG-----ACAGTCCGAGAAACAAAGGACCATGGATGCTGCTGCTGC 1358  
QY 540 GCGAGAGCTATCCCACTGGGACTTCCGAGGCACTTGAACCTCAGAACACTACAGCGG 599  
Db 1359 GAGAGACGATATCCCAACTGGATTCTAAGCAGCTTAACCTCAGAACACTACGCGCA 1418  
QY 600 AGACGCCACCGGTGCTTG 618  
Db 1419 AGAGACCGCGGGTCTTG 1437  
  
RESULT 14  
AC022301  
LOCUS 187868 bp DNA HTG 15-JUN-2000  
DEFINITION Mus musculus clone RP23-329K24, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 59  
unordered pieces.  
ACCESSION AC022301  
VERSION AC022301.9 GI:8493517  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 187868)  
Molecular, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,  
Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T.,  
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,  
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,  
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,  
Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,  
Gorell, J. H., Gunaratne, P., Haller, G., Hernandez, J., Hogues, M.,  
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,  
Kovar, C., Liu, J., Liu, W., Louised, H., Lozada, R. J., Martin, R.,  
Massey, E., McLeod, M. P., Mei, G., Moore, S., Morgan, M., Morris, S.,  
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogihara, M., Parish, B.,  
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,  
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,  
Muzny, D. M., Rives, M., Scherer, S., Sodergren, E., Weinstein, G.,  
Worley, K. and Gibbs, R.  
Direct Submission  
Unpublished  
2 (bases 1 to 187868)  
Worley, K. C.  
Direct Submission  
Submitted (29-JAN-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jun 13, 2000 this sequence version replaced gi:8248590.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
-----  
Center project name: MAAX  
Center clone name: RP23-329K24  
----- Summary Statistics  
-----  
Sequencing vector: M13; L08021

Chemistry: Dye-primer Bodipy: 82% of reads  
Chemistry: Dye-terminator Big Dye: 18% of reads  
Assembly program: Phrap: version 0.990329  
Consensus quality: 13377 bases at least Q40  
Consensus quality: 159061 bases at least Q30  
Consensus quality: 169688 bases at least Q20  
Estimated insert size: 166938; sum-of-contigs estimation  
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation

NOTE: This is a 'working draft' sequence. It currently  
\* consists of 59 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 7154: contig of 7154 bp in length  
7155 7254: gap of unknown length  
7255 14340: contig of 7086 bp in length  
14341 14440: gap of unknown length  
14441 22916: contig of 8476 bp in length  
22917 23016: gap of unknown length  
23017 30834: contig of 7818 bp in length  
30835 30934: gap of unknown length  
30935 37993: contig of 7059 bp in length  
37994 38093: gap of unknown length  
38094 45511: contig of 7418 bp in length  
45512 45611: gap of unknown length  
45612 51842: contig of 6231 bp in length  
51843 51942: gap of unknown length  
51943 58572: contig of 6630 bp in length  
58573 58672: gap of unknown length  
58673 62718: contig of 4046 bp in length  
62719 67974: gap of unknown length  
67975 68074: contig of 5136 bp in length  
68075 72883: contig of 4809 bp in length  
72884 72984: gap of unknown length  
72985 76727: contig of 3743 bp in length  
76728 76826: gap of unknown length  
76827 79331: contig of 2805 bp in length  
79332 79731: gap of unknown length  
79732 82858: contig of 3127 bp in length  
82859 82958: gap of unknown length  
82959 86054: contig of 3096 bp in length  
86055 86154: gap of unknown length  
86155 90283: contig of 4129 bp in length  
90284 90383: gap of unknown length  
90384 94598: contig of 4215 bp in length  
94599 99014: gap of unknown length  
99015 99114: contig of 4316 bp in length  
99115 103295: contig of 4181 bp in length  
103296 103395: gap of unknown length  
103396 106698: contig of 3303 bp in length  
106699 106798: gap of unknown length  
106799 106957: contig of 2859 bp in length  
106958 109757: gap of unknown length  
109758 113392: contig of 3635 bp in length  
113393 113492: gap of unknown length  
113493 117011: contig of 3519 bp in length  
117012 117111: gap of unknown length  
117112 120036: contig of 2925 bp in length  
120037 120136: gap of unknown length  
120137 122449: contig of 2313 bp in length  
122450 122549: gap of unknown length  
122550 124756: contig of 1726 bp in length  
124757 124776: gap of unknown length  
124777 126297: contig of 1921 bp in length  
126298 126397: gap of unknown length  
126398 129180: contig of 2784 bp in length  
129181 129280: gap of unknown length

BASE COUNT 66 a 117 c 61 g 52 t 1 others  
ORIGIN

/db\_xref="taxon:32644"

Query Match 12.3%; Score 151; DB 81; Length 297;  
Best Local Similarity 76.0%; Pred. No. 1.6e-24;  
Matches 212; Conservative 0; Mismatches 65; Indels 2; Gaps 2;  
Qy 1 ATGTGTCACTCTGCGAGCTGCCACCGACCATGACCATCTGCGAGGCCCGGACCCCGGCC 60  
Db 16 ATGGAACACTCTGCGGCTGCCACTCCCCATGACCACTACAGGCCCTAACCCGAGCC 75  
Qy 61 CCTTCCACCATGCCGGGACCGCGGGGTCCGGTCTTGAGATCTTACCTTCGACCCCT 120  
Db 76 CCTCACGTTATCGTCCAGCCACAATA-GCTGCGGCCCTGAGATCTCCACCTCAGACCCCT 134  
Qy 121 CTCCCGAGCGCGAGCGGCCCTGCGCGGCGGCCCGCCAGCCCTCTCTCGCGGGCACCGAAG 180  
Db 135 CTCCGGAAGCCACAGCGGCCCTGCGCGGCGGCCCGCCAGCTTC-CTCGAGGCCACCAAG 193  
Qy 181 CGCAGCGCGAGGGTCTCTACCTCGAGTGTGCGGCGCGAGCTGCCAGTCGAGGACCG 240  
Db 194 CGCAGTTCAGGAATCTCTACCTAGAGTGTGCGGCGCGAGTTCGCCAGTTAAGGGACCA 253  
Qy 241 AACCCAGCCAAAGGCTTCTTCTTCTGCTGCTCACCATC 279  
Db 254 AACCTGCCAGAGGGCTTCTTTTCGGGTGATCGCCATC 292

Search completed: January 30, 2001, 19:23:39  
Job time: 25905 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2001, 19:23:39 ; Search time 1900.13 seconds  
(without alignments)  
694.889 Million cell updates/sec

Title: US-08-799-910-9\_COPY\_211\_468

Perfect score: 258

Sequence: 1 GTCCGGCCAGCTCCAGT.....AGCAACACCGCGCGCCTTC 258

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 255875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_ph:\*  
6: gb\_pl1:\*  
7: gb\_pl2:\*  
8: gb\_pri:\*  
9: gb\_pr2:\*  
10: gb\_ro:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_un:\*  
14: em\_fun:\*  
15: em\_hum1:\*  
16: em\_hum2:\*  
17: em\_in:\*  
18: em\_om:\*  
19: em\_or:\*  
20: em\_ov:\*  
21: em\_pat:\*  
22: em\_ph:\*  
23: em\_pi:\*  
24: em\_ro:\*  
25: em\_sts:\*  
26: em\_sy:\*  
27: em\_un:\*  
28: em\_v1:\*  
29: gb\_ba3:\*  
30: gb\_in1:\*  
31: gb\_in2:\*  
32: gb\_in3:\*  
33: gb\_pi3:\*  
34: gb\_pr4:\*  
35: em\_ba1:\*  
36: em\_ba2:\*  
37: em\_ba3:\*  
38: em\_ba4:\*  
39: em\_ba5:\*  
40: em\_ba6:\*  
41: em\_ba7:\*  
42: em\_ba8:\*  
43: em\_ba9:\*

44: em\_htg8:\*  
45: em\_htg9:\*  
46: em\_htg10:\*  
47: em\_hum3:\*  
48: em\_hum4:\*  
49: em\_hum5:\*  
50: em\_hum6:\*  
51: gb\_pr5:\*  
52: gb\_pr6:\*  
53: gb\_pr7:\*  
54: gb\_htg1:\*  
55: gb\_htg2:\*  
56: gb\_htg3:\*  
57: gb\_htg4:\*  
58: gb\_htg5:\*  
59: gb\_htg6:\*  
60: gb\_htg7:\*  
61: gb\_htg8:\*  
62: gb\_htg9:\*  
63: gb\_htg10:\*  
64: gb\_htg11:\*  
65: gb\_htg12:\*  
66: gb\_htg13:\*  
67: gb\_htg14:\*  
68: gb\_htg15:\*  
69: gb\_htg16:\*  
70: gb\_htg17:\*  
71: gb\_htg18:\*  
72: gb\_htg19:\*  
73: gb\_htg20:\*  
74: gb\_htg21:\*  
75: gb\_htg22:\*  
76: gb\_htg23:\*  
77: gb\_sts1:\*  
78: gb\_sts2:\*  
79: gb\_v11:\*  
80: gb\_v12:\*  
81: gb\_pat1:\*  
82: gb\_pat2:\*  
83: em\_htg0:\*  
84: gb\_htg24:\*  
85: gb\_pr8:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	258	100.0	477	10	AF083421	AF083421 Homo sapi
2	258	100.0	1230	53	HSDF2	Y14551 Homo sapien
3	258	100.0	1309	10	AF039067	AF039067 Homo sapi
4	258	100.0	1693	10	AF071596	AF071596 Homo sapi
5	258	100.0	1864	53	HSPRG1	X96438 H.sapiens P
6	258	100.0	44118	9	AC006165	AC006165 Homo sapi
7	258	100.0	90244	8	AB023051	AB023051 Homo sapi
8	258	100.0	200000	51	AP000512	AP000512 Homo sapi
9	256.4	99.4	1223	85	S81914	S81914 IEX-1-radia
10	154.4	59.8	1758	11	RNPRG1	X96437 R.norvegicu
11	151.2	58.6	203476	70	AC074150	AC074150 Mus muscu
12	149.6	58.0	1938	11	MGGLY96	X67644 M.musculus
13	139.2	54.0	187868	62	AC022301	AC022301 Mus muscu
14	44	17.1	297	81	A75448	A75448 Sequence 11
15	44	17.1	297	81	A78427	A78427 Sequence 11
16	43.2	16.7	115468	57	AC011820	AC011820 Homo sapi
17	43	16.7	7218	81	I66494	I66494 Sequence 14
18	40.2	15.6	34725	29	SC1B2	AL358812 Streptomy
19	39.8	15.4	3897	2	MXU40656	U40656 Mycobacteri
20	39.6	15.3	33818	2	MTCY159	283863 Mycobacteri
21	39.4	15.3	2003	85	HUMDAFC1	M64356 Human decay

```

/db_xref="taxon:9606"
/cell_type="leukocytes"
<1..11
/gene="dif-2"
1..1230
/gene="dif-2"
12..482
/gene="dif-2"
/codon_start=1
/product="DIF-2 protein"
/protein_id="CAA74886.1"
/db_xref="GI:2660551"
/db_xref="SWISS-PROT:P46695"
/translation="MCHSRSCPTMTILQAPTAPSTIPGPRKSGPETFTFDLPPEP
AAPAGRPASGRHKRRSRVLYPRVRRQLPVEENPAKRLLELLLTIVFQCILMAE
EGVAPALPPEDAPNAASLAPTPVSPVLEPNTSEPSYALDLSFLQOHPAAF"
172
/gene="dif-2"
/citation={3}
/replace="g"
328
/gene="dif-2"
/citation={3}
/replace="g"
483..1230
/gene="dif-2"
673..675
/gene="dif-2"
/citation={3}
/replace="ag"
687..689
/gene="dif-2"
/citation={3}
/replace="tg"
783..785
/gene="dif-2"
/citation={3}
/replace="tg"
1020..1021
/gene="dif-2"
/citation={3}
/replace="att"
1120
/gene="dif-2"
/citation={3}
/replace="g"
1133
/gene="dif-2"
/citation={3}
/replace="t"

BASE COUNT 259 a 367 c 342 g 262 t
ORIGIN

Query Match 100.0%; Score 258; DB 53; Length 1230;
Best Local Similarity 100.0%; Pred. No. 6.8e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGCGCCAGTGCAGTGCAGGAACCGAACCCAGCCGCTTCTTTCTGCTG 60
DB 222 GTCGGCGCCAGTGCAGTGCAGGAACCGAACCCAGCCGCTTCTTTCTGCTG 281
QY 61 CTCACCATCGTCTTCTGCGAGATCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 120
DB 282 CTCACCATCGTCTTCTGCGAGATCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 341
QY 121 CCAGAGGAGCCCTTAACCCGATCCCTGCGGCGCCACCCCTGCTGCTGCTGCTG 180
DB 342 CCAGAGGAGCCCTTAACCCGATCCCTGCGGCGCCACCCCTGCTGCTGCTGCTG 401
QY 181 CCTTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCTCCAG 240
DB 402 CCTTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGACACTTCTCTCCAG 461

```

```

QY 241 CAACACCGCGCCCTTC 258
DB 462 CAACACCGCGCCCTTC 479

RESULT 3
AF039067 1309 bp mRNA PRI 21-AUG-1998
LOCUS Homo sapiens anti-death protein (IEX-1L) mRNA, complete cds.
DEFINITION AF039067
VERSION AF039067.1 GI:3449375
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Wu.M.X., Ao.Z., Prasad.K.V.S., Wu.R. and Schlossman.S.F.
TITLE IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell
survival
JOURNAL Science 281 (5379), 998-1001 (1998)
MEDLINE 98369175
REFERENCE 2 (bases 1 to 1309)
AUTHORS Ao.Z. and Wu.M.X.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1997) Tumor Immunology, Dana Farber Cancer
Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES
Location/Qualifiers
1..1309
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1309
/gene="IEX-1L"
1..582
/gene="IEX-1L"
/feature="type II membrane protein"
/codon_start=1
/product="anti-death protein"
/protein_id="AAC32558.1"
/db_xref="GI:3449376"
/translation="MCHSRSCPTMTILQAPTAPSTIPGPRKSGPETFTFDLPPEP
AAPAGRPASGRHKRRSRVLYPRVRRQLPVEENPAKRLLELLLTIVFQCILMAE
FCLPLVRRQLPVEENPAKRLLELLLTIVFQCILMAE
PVSPVLEPNTSEPSYALDLSFLQOHPAAF"
BASE COUNT 262 a 393 c 369 g 285 t
ORIGIN

```

```

Query Match 100.0%; Score 258; DB 10; Length 1309;
Best Local Similarity 100.0%; Pred. No. 6.7e-47;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGCGCCAGTGCAGTGCAGGAACCGAACCCAGCCGCTTCTTTCTGCTG 60
DB 322 GTCGGCGCCAGTGCAGTGCAGGAACCGAACCCAGCCGCTTCTTTCTGCTG 381
QY 61 CTCACCATCGTCTTCTGCGAGATCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 120
DB 382 CTCACCATCGTCTTCTGCGAGATCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCCT 441
QY 121 CCAGAGGAGCCCTTAACCGGCGATCTTGGCGCCACCCCTGCTGCTGCTGCTG 180
DB 442 CCAGAGGAGCCCTTAACCGGCGATCTTGGCGCCACCCCTGCTGCTGCTGCTG 501
QY 181 CCTTTTAATCTGACTTCGGAGCCCTTCGGACTACGCTCTGGACCTCAGACACTTCTCTCCAG 240
DB 502 CCTTTTAATCTGACTTCGGAGCCCTTCGGACTACGCTCTGGACCTCAGACACTTCTCTCCAG 561
QY 241 CAACACCGCGCCCTTC 258
DB 562 CAACACCGCGCCCTTC 579

```

Matches 258: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCCGAGCTGCCAGTCGAGAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
 (|||||) (|||||) (|||||) (|||||) (|||||) (|||||) (|||||) (|||||) (|||||) (|||||)  
 Db 914 GTCCGGCCGAGCTGCCAGTCGAGAACCCAGCCAAAGGCTTCTTTCTGCTG 973  
 QY 61 CTCACCATCGTCTTCTCCGATCCTGATGGCTGAAGAGGTGTGCGGGCCGCTGCCT 120  
 Db 974 CTCACCATCGTCTTCTCCGATCCTGATGGCTGAAGAGGTGTGCGGGCCGCTGCCT 1033  
 QY 121 CCAGAGAGCCCTATAGCGGGCATCCTGGCGCCACCCCTGTGTGCCCGTCTCTGAG 180  
 Db 1034 CCAGAGAGCCCTATAGCGGGCATCCTGGCGCCACCCCTGTGTGCCCGTCTCTGAG 1093  
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGACCTCAGCACTTCTCCACG 240  
 Db 1014 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGACCTCAGCACTTCTCCACG 1153  
 QY 241 CAACACCGCGCGCTTC 258  
 Db 1154 CAACACCGCGCGCTTC 1171

## RESULT 6

AC006165 44118 bp DNA PRI 08-DEC-1998  
 LOCUS Homo sapiens clone UMG:Y54C125 from 6p21, complete sequence.  
 AC006165  
 AC006165.1 GI:3980464  
 HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 44118)  
 AUTHORS Janer, M., Guillaudoux, T., Vu, Q., Kutayavin, T., Harter, H. and Geraghty, D.E.  
 TITLE Large scale sequence analysis of the human MHC class I region  
 JOURNAL Unpublished (1998)  
 REMARK Fred Hutchinson Cancer Research Center  
 The Clinical Research Division  
 1100 Fairview Ave. N., P.O. Box 19024  
 Seattle, WA 98109-1024  
 REFERENCE 2 (bases 1 to 44118)  
 AUTHORS Geraghty, D.E. and Olson, M.V.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-DEC-1998) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA  
 REMARK University of Washington Human Genome Center  
 Box 352145, Seattle, WA 98195  
 Contact: Daniel E. Geraghty (geraghty@fhcr.org)  
 Overlapping Sequences:  
 5': UMG:Y54C283  
 3': UMG:Y54C222 (Genbank Accession: AC006049)  
 -----  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.  
 -----  
 Double stranded (DS) coverage: 58.8%  
 DS or two chemistry coverage: 99.5%  
 Single stranded regions: 2  
 -----  
 Sequence Validation:  
 This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below.

Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

Map	Seq	HindIII	Map	Seq	NsiI
6714..50	6452..00	1823..23	1794..00	5581..12	5401..00
1089..04	1085..00	1246..04	1228..00	10570..24	10123..00
1536..15	1528..00	5570..33	5372..00	10069..29	9626..00
3623..08	3537..00	1670..92	1630..00		
3784..08	3621..00	1270..76	1280..00		
6834..00	6689..00	4235..79	4109..00		
11234..36	10874..00	1171..00	1151..00		
762..43	762..00	3226..11	3096..00		
2727..38	2653..00	10501..64	10080..00		
1691..42	1684..00	4208..44	4052..00		
1295..60	1277..00				
746..87	734..00				
-----					
FEATURES		Location/Qualifiers			
Source		1..44118			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/chromosome="6"			
		/map="6p21"			
		/clone="CGM1.D100E8"			
		/sub_clone="UMG:Y54C125"			
		/cell_line="CGM1"			
		/clone_lib="Wash U YAC library"			
		218..405			
repeat_region		/rpt_family="MER3"			
		400..999			
repeat_region		/rpt_family="Alu"			
		1797..2378			
repeat_region		/rpt_family="Alu"			
		2649..2956			
repeat_region		/rpt_family="Alu"			
		complement(3372..3894)			
repeat_region		/rpt_family="Alu"			
		3941..4070			
repeat_region		/rpt_family="Alu"			
		complement(4449..4724)			
repeat_region		/rpt_family="Alu"			
		complement(5117..5389)			
repeat_region		/rpt_family="Alu"			
		7668..7913			
repeat_region		/rpt_family="Alu"			
		complement(7924..8226)			
repeat_region		/rpt_family="Alu"			
		8250..8551			
repeat_region		/rpt_family="Alu"			
		complement(8893..8973)			
repeat_region		/rpt_family="MIR"			
		complement(11829..12096)			
repeat_region		/rpt_family="Alu"			
		12416..12740			
repeat_region		/rpt_family="Alu"			
		complement(13181..13247)			
repeat_region		/rpt_family="Alu"			
		complement(13498..13647)			



## COMMENT

-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----

Project Information  
Center Project Name: 0  
Center clone name: RPCI-23\_128D3  
-----

Summary Statistics  
Consensus quality: 139878 bases at least Q40  
Consensus quality: 165603 bases at least Q30  
Consensus quality: 177139 bases at least Q20  
Estimated insert size: 183300; agarose-fp estimation  
Estimated insert size: 200576; sum-of-contigs estimation  
Quality coverage: 3.48 in Q20 bases; agarose-fp estimation  
Quality coverage: 3.48 in Q20 bases; sum-of-contigs estimation  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 30 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1175: contig of 1175 bp in length  
1176: gap of unknown length  
1276: contig of 1288 bp in length  
2563: gap of unknown length  
2564: gap of unknown length  
3908: contig of 1243 bp in length  
4008: gap of unknown length  
5224: contig of 1218 bp in length  
5225: gap of unknown length  
5325: contig of 1312 bp in length  
6636: gap of unknown length  
6737: contig of 2460 bp in length  
9197: gap of unknown length  
9297: contig of 1427 bp in length  
10724: gap of unknown length  
10824: contig of 2029 bp in length  
12853: gap of unknown length  
12953: contig of 3073 bp in length  
16026: gap of unknown length  
16126: gap of unknown length  
16126: contig of 2971 bp in length  
19097: gap of unknown length  
19196: contig of 2554 bp in length  
21750: gap of unknown length  
21850: contig of 4729 bp in length  
21851: gap of unknown length  
26579: gap of unknown length  
26580: gap of unknown length  
32472: contig of 5793 bp in length  
32473: gap of unknown length  
32573: contig of 4422 bp in length  
36994: gap of unknown length  
37094: contig of 6977 bp in length  
44071: gap of unknown length  
44171: gap of unknown length  
50721: contig of 6556 bp in length  
50827: gap of unknown length  
50828: contig of 6377 bp in length  
57204: gap of unknown length  
57304: contig of 7772 bp in length  
65076: gap of unknown length  
65176: gap of unknown length  
71824: contig of 6648 bp in length  
71924: gap of unknown length  
78035: contig of 6111 bp in length  
78135: gap of unknown length  
78136: contig of 9703 bp in length  
87838: gap of unknown length  
87939: contig of 6986 bp in length  
94924: gap of unknown length  
95024: gap of unknown length  
101189: contig of 6164 bp in length  
101289: gap of unknown length  
108925: contig of 7637 bp in length  
108926: gap of unknown length

\* 109026 119738: contig of 10713 bp in length  
\* 119739: gap of unknown length  
\* 119839: contig of 10038 bp in length  
\* 129876: gap of unknown length  
\* 129877: contig of 13267 bp in length  
\* 143243: gap of unknown length  
\* 143244: contig of 17821 bp in length  
\* 143344: gap of unknown length  
\* 161165: contig of 19032 bp in length  
\* 161265: gap of unknown length  
\* 180317: contig of 23060 bp in length  
\* 180417: contig of 23060 bp in length.

FEATURES  
Location/Qualifiers  
1..203476  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-128D3"  
/clone\_lib="RPCI mouse BAC library 23"

BASE COUNT 51748 a 48714 c 48138 g 51580 t 3296 others  
ORIGIN

Query Match 58.6%; Score 151.2; DB 70; Length 203476;  
Best Local Similarity 75.9%; Pred. No. 4.7e-24;  
Matches 205; Conservative 0; Mismatches 53; Indels 12; Gaps 1;

OY 1 GTCGGCGCCAGCTGCCAGTCGAGGAACCGAACCCAGCAAGGCTCTCTTTCTGCTG 60  
|||||  
Db 92875 GTCGGCGCCAGCTTACCACCGAGGAACCAACATTCGCAAGAGGCTCTTTCTGCTG 92934  
OY 61 CTCACCATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGCTGTGCGGGGCGCCCTGCCT 120  
|||||  
Db 92935 TTGCCATCATCTTCTGCCAGATTTTTCATGGCTCAAGAGGCTGTGTCGACGCCCTGGCT 92994  
OY 121 CCAGAGACGCCCTTACGCGCGCATCC-----TGGCGCCACCCCTGTGTCC 168  
|||||  
Db 92995 CCGAGGATGCTTACCAGCGCGGTGACACCTGAGCCCATTTCTGCGCCCATTAATCTGCGCCC 93054  
OY 169 CCCTCTCGAGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGC 228  
|||||  
Db 93055 CCGTCTCGAGCCCTTGAACCTGACCTCGAGTCTCGGAGTCTCGGACTATGCGTGGATCTTAAA 93114  
OY 229 ACTTCTCCAGCAACACCGCGCGCCTTC 258  
|||||  
Db 93115 GCTTTCTCCAGCAACATCGCGCGCCTTC 93144

## RESULT 12

MMGLY96

LOCUS 1938 bp mRNA ROD 11-MAR-1993  
DEFINITION M.musculus gly96 mRNA.  
ACCESSION X67644  
VERSION X67644.1 GI:287803  
KEYWORDS gly96 gene; glycosylated protein.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 1938)  
AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
TITLE Direct Submission  
JOURNAL Submitted (29-JUL-1992) L. Lau, University of Illinois College of  
Med., Dept. of Genetics, M/C 669, 808 South Wood Street, Chicago,  
IL 60612, USA

## REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

1..1938

/organism="Mus musculus"

Location/Qualifiers  
1..1938  
/organism="Mus musculus"

Charles C. H. Yoon, J. K., Sinske, J. S. and Lau, L. F.

Genomic structure; cDNA sequence, and expression of gly96, a growth  
factor-induced immediate-early gene encoding a short-lived  
glycosylated protein

Oncogene 8 (3), 797-801 (1993)

93173526

Location/Qualifiers

1..1938

/organism="Mus musculus"

